Decisional-Emotional Support System for a Synthetic Agent – Influence of Emotions in Decision-Making Toward the Participation of Automata in Society

Javier Francisco Guerrero Rázuri
Decisional-Emotional Support System for a Synthetic Agent

Influence of Emotions in Decision-Making Toward the Participation of Automata in Society

Javier Francisco Guerrero Rázuri
As the seaman rejoices
When his ship reaches the port,
So does the writer rejoice
When he writes the final word.
Abstract

Emotion influences our actions, and this means that emotion has subjective decision value. Emotions, properly interpreted and understood, of those affected by decisions provide feedback to actions and, as such, serve as a basis for decisions. Accordingly, “affective computing” represents a wide range of technological opportunities toward the implementation of emotions to improve human-computer interaction, which also includes insights across a range of contexts of computational sciences into how we can design computer systems to communicate and recognize the emotional states provided by humans. Today, emotional systems such as software-only agents and embodied robots seem to improve every day at managing large volumes of information, and they remain emotionally incapable to read our feelings and react according to them. From a computational viewpoint, technology has made significant steps in determining how an emotional behavior model could be built; such a model is intended to be used for the purpose of intelligent assistance and support to humans. Human emotions are engines that allow people to generate useful responses to the current situation, taking into account the emotional states of others. Recovering the emotional cues emanating from the natural behavior of humans such as facial expressions and bodily kinetics could help to develop systems that allow recognition, interpretation, processing, simulation, and basing decisions on human emotions. Currently, there is a need to create emotional systems able to develop an emotional bond with users, reacting emotionally to encountered situations with the ability to help, assisting users to make their daily life easier. Handling emotions and their influence on decisions can improve the human-machine communication with a wider vision. The present thesis strives to provide an emotional architecture applicable for an agent, based on a group of decision-making models influenced by external emotional information provided by humans, acquired through a group of classification techniques from machine learning algorithms. The system can form positive bonds with the people it encounters when proceeding according to their emotional behavior. The agent embodied in the emotional architecture will interact with a user, facilitating their adoption in application areas such as caregiving to provide emo-
tional support to the elderly. The agent’s architecture uses an adversarial structure based on an Adversarial Risk Analysis framework with a decision analytic flavor that includes models forecasting a human’s behavior and their impact on the surrounding environment. The agent perceives its environment and the actions performed by an individual, which constitute the resources needed to execute the agent’s decision during the interaction. The agent’s decision that is carried out from the adversarial structure is also affected by the information of emotional states provided by a classifiers-ensemble system, giving rise to a “decision with emotional connotation” included in the group of affective decisions. The performance of different well-known classifiers was compared in order to select the best result and build the ensemble system, based on feature selection methods that were introduced to predict the emotion. These methods are based on facial expression, bodily gestures, and speech, with satisfactory accuracy long before the final system.
This thesis is dedicated to my beloved parents and in loving memory of my uncle Beto, who now rests in heaven.
List of Papers

The following papers, referred to in the text by their Roman numerals, are included in this thesis.

PAPER I: An Adversarial Risk Analysis Model for an Emotional Based Decision Agent

PAPER II: An adversarial risk analysis model for an autonomous imperfect decision agent

PAPER III: An Adversarial Risk Analysis Model for a Decision Agent facing Multiple Users

PAPER IV: Automatic Emotion Recognition through Facial Expression Analysis in Merged Images Based on an Artificial Neural Network

PAPER V: Effect of emotional feedback in a decision-making system for an autonomous agent

PAPER VI: Recognition of emotions by the emotional feedback through behavioral human poses

PAPER VII: Speech emotion recognition in emotional feedback for Human-Robot Interaction

PAPER VIII: Decision-making content of an agent affected by emotional feedback provided by capture of human’s emotions through a Bimodal System

Reprints were made with permission from the publishers.
Author’s contribution

The major contributions of this dissertation that will be presented to the research community are summarized in a group of publications as follows:

Publication I  *An Adversarial Risk Analysis Model for an Emotional Based Decision Agent* (I)

This publication was the first attempt to provide an idea of a decisional model focused on an autonomous agent. The approach is supported over the first scheme of a mathematical model using the Adversarial Risk Analysis (ARA) framework. The model tries to generate behavior patterns of forecasting actions of an agent, taking into account the analysis of the context in which risks stem from deliberate actions of intelligent adversaries. The first decisional model of an agent introduces the first three elements that establish among themselves the variables that will interact with humans within an environment. The author’s main contribution was the development of the first scheme of a model to control the behavior of an agent facing another intelligent adversary that could be a human. The model has a core of a multi-attribute decision framework with the support of forecasting models of the adversary (Adversarial Risk Analysis) and emotion-based behavior (Affective Decision Making).

Publication II  *An Adversarial Risk Analysis Model for an Autonomous Imperfect Decision Agent* (II)

The author’s main contribution in this publication was focused on the continuity of the multi-attribute decision analysis model, complemented by forecasting models of the adversary. To investigate selected parts of the full model and their proper operation, we have performed tests over the first prototype simulator from agent behavior, developed in a C++ integrated development environment based on the Eclipse platform, with only a few human and agent actions. We assume that our agent resides within an environment in which there is another
person with whom it interacts. The simulator was based in a real edutainment robot that is endowed with a sensorimotor system (including vision, audition, touch sensors, temperature sensor, and inclination sensor). Using these sensors, it may capture information about its environment, other users, and other robots, which it interprets to infer the other agents’ actions and the state of the environment. Such information is evaluated to determine its impact on the agent’s objectives, and to learn to evolve to the environment and make forecasts based on the mathematical models.

Publication III  *An Adversarial Risk Analysis Model for a Decision Agent Facing Multiple Users* (III)
This publication addresses the extension of the support decision-making model of the autonomous agent to several individuals whom it might encounter. The modification was also intended to extend a case in which there are several agents, possibly cooperating or competing, depending on their emotional states. We also open the possibility of learning about new user actions, based on repeated readings, augmenting the number of individuals within the interaction Loop. The author contributed with the modification of the model and its intended future use, taking into account that the forecasts and valuations can suffer the influence of external emotions provided by humans and the impact on the decision-making process. This appreciation should also keep open the possibility of model flexibility.

Publication IV  *Automatic Emotion Recognition through Facial Expression Analysis in Merged Images Based on an Artificial Neural Network* (IV)
This publication addresses one of the first attempts from the group of methods in automatic emotion recognition included in this dissertation related to recognizing human emotion through facial gestures, visualizing the future use of the emotional cues in the decisional system. The author’s main contribution was to introduce an approach to analyze two areas of the human face, the zone of the eyes and the mouth, in order to merge them into a single new image using a classification technique of emotional information. The eyes and mouth are taken as the basis for the decryption of emotion because of the predictability and visibility of recurrent emotional changes in
those areas. The machine learning method implemented is a feed-forward neural network trained by back-propagation that uses as input pixels from each processed image as emotional features. The simulations in MATLAB provided information related to the accuracy of the proposed algorithm to detect emotional information by facial gestures.

Publication V  Effect of Emotional Feedback in a Decision-Making System for an Autonomous Agent

In this publication, the author contributes a new concept henceforth termed “emotional feedback” that reinforces the benefits on the closed-loop human-agent interaction framework. This publication highlighted the idea to combine the decision-making model with the capability to use emotional inputs from human facial gestures in order to generate an emotionally intelligent decision in an agent. This implementation established the initial foundations to modify the agent’s behavior. The decision agent will be able to make decisions influenced by the emotional information conveyed through the human face, using the framework of Adversarial Risk Analysis (ARA) and the machine learning method of feed-forward neural network to decrypt the emotional state. Here, the author showed that such an approach demonstrates feasibility in handling the agent’s behavior in a positive way, achieving self-regulation between actions and emotions. A new paradigm of multi-disciplinary research is introduced, as reflected in the support of the broaden-and-build theory, for the exploration of the evolving function of positive emotions of the agent. In this case, the agent’s reactions when it is facing the opponent’s positive behavior could trigger a cycle of positive emotions connected to the relevant action. Here, the author presented several results using a new version of the prototype simulator from the agent behavior, developed in the object-oriented programming language Python. The publication further examines the simulations that cover the evolution of the expected utility just one period ahead, the agent’s reactive behavior facing the opponent’s actions through the weighting of posterior model probability, and the evolution of the “emotional feedback” provided by the group of facial gestures from humans and the evolution of the agent’s behavior using the established rule-based design of emotional feedback for positive
evolution of the agent’s behavior.

Publication VI  *Recognition of Emotions by the Emotional Feedback through Behavioral Human Poses (VI)*

This publication contributes with an approach for the decryption of emotions through bodily expression, leaving the possibility open to use the technique to acquire emotional intelligence in a machine without consuming the large amounts of memory that other techniques require. This publication states the relationship between dynamic postures from humans and attributions of emotion in an attempt to describe how emotion may be communicated through the body. The approach proposed is related to the analysis and processing of images from a human skeleton acquired by the sensing input device of a Microsoft Xbox 360. Several machine learning algorithms from the Waikato Environment for Knowledge Analysis (WEKA) were tested to achieve an accuracy classification, taking into account that the emotional outputs could be used in the decisional model of the agent. The approach reveals that the analysis of images from human body poses makes it possible to obtain relevant information through the combination of proper data in the same image. In addition, we proposed the use of a vector of features named *matrix-knowledge* that was prepared for the analysis with the several machine learning techniques, in which the sum of the images represents the knowledge of the kinetic behavior of an emotional pose.

Publication VII  *Speech Emotion Recognition in Emotional Feedback for Human-Robot Interaction (VII)*

This publication describes an approach to decrypt emotional information from human speech cues. The approach uses speech cues normally described in Music Information Retrieval (MIR) for automatic speech recognition. The most commonly used features in several studies to capture emotional characteristics in time and frequency were selected. All of the information provided for several audio samples with emotional content were collected in several vectors useful to construct an emotional speech database. Here, the author showed that the feature selection of emotional cues allows the reduction of dimensionality of the data, in turn leading to fewer computational processes, which facilitates a low-cost
robotic platform. The author’s main contribution was the development of techniques to perform the parameterization of audio data for recognizing emotion in speech. The techniques used are grouped in processing stages of audio data to build the feature vector and to use machine learning classifiers from the Waikato Environment for Knowledge Analysis (WEKA) for discrimination of emotional cues.

Publication VIII *Decision-Making Content of an Agent Affected by Emotional Feedback Provided by Capture of Human’s Emotions through a Bimodal System* [VIII]

This publication describes the entire model as a fusion of the decisional and the emotional recognition stages, using facial gestures and speech as emotional sources, the concept of “emotional feedback” from the broaden-and-build theory, and the first two stages of emotional self-regulation in order to make the decision process of the agent more spontaneous, with long-term best interest considered. This publication shows that emotions emerge as a key concept that has a major impact in communication and decision-making as main components of intelligent behavior in a machine also known as a synthetic agent. The author’s main contribution is the proposal of a fusion between two emotional sources that gave rise to the creation of a bimodal system composed of two baseline unimodal classifiers and a multi-classifier. The proposed emotional acquisition stage consisted of preprocessing, feature extraction, and pattern classification steps. Preprocessing and feature extraction methods were devised so that emotion-specific characteristics could be extracted in an audio-visual scheme. Within the proposed interaction system, we can consider that the final decision of the agent might still be an affective decision, taking into account that the agent predicts the affective consequences of each available alternative to not disturbing the individual that it encounters, and this entails a strengthening of relations between human and machine within an emotional loop.
# Contents

Abstract \[v\]

List of Papers \[ix\]

Author’s contribution \[xi\]

Abbreviations \[xxi\]

List of Figures \[xxiii\]

List of Tables \[xxv\]

Acknowledgements \[xxvii\]

## 1 Introduction

1.1 Motivation \[31\]

1.2 Problem Statement \[33\]

1.3 Research Question \[37\]

1.4 Research Methodology \[39\]

1.4.1 Philosophical Assumptions \[40\]

1.4.2 The Design Science Paradigm \[43\]

1.4.3 Design Science and the Scientific Enquiry \[44\]

1.4.4 Contributions \[46\]

1.4.5 Dissertation Structure \[49\]

1.5 Summary \[50\]

## 2 Related Work

2.1 Affective Computing \[53\]

2.2 The Understanding and Usefulness of Human Emotions within the AI Community \[55\]

2.3 Required Backing to Pursuit of a “Decision with Emotional Connotation” \[61\]

2.3.1 Adversarial Risk Analysis \[61\]
# Table of Contents

2.3.2 Emotion Recognition by Machine Learning ........................................ 64
2.4 Classifier Methods and Performance Measures ...................................... 73
   2.4.1 Classifier Methods ........................................................................ 73
   2.4.2 Performance Measures .................................................................... 74
   2.4.3 Multi-classifiers ............................................................................. 76
2.5 Summary ................................................................................................. 77

3 Decision-Making Model of an Agent .......................................................... 79
   3.1 Basic Framework .................................................................................. 79
   3.2 Adversarial Risk Analysis Decision Model ......................................... 80
      3.2.1 Forecasting Models ...................................................................... 80
      3.2.2 Preference Model ......................................................................... 81
      3.2.3 Expected Utility ........................................................................... 81
   3.3 Implementation ..................................................................................... 82
      3.3.1 Basic Elements ............................................................................ 82
      3.3.2 Forecasting Models ...................................................................... 84
      3.3.3 Multiobjective Preference Model ................................................ 87
      3.3.4 Results and Discussion ................................................................ 89
   3.4 Summary ............................................................................................... 90

4 Emotional Acquisition Through Facial Gestures ........................................... 97
   4.1 Implementation of Facial Emotion Detection Model .............................. 98
   4.2 Results and Discussion ....................................................................... 102
   4.3 Summary ............................................................................................. 105

5 Decision-Making Model Affected by Emotional Feedback .......................... 107
   5.1 Scenario Focused on the Interaction ..................................................... 107
      5.1.1 Rule-Based Design for the Emotional Feedback ............................ 109
      5.1.2 Results and Discussion ................................................................ 110
   5.2 Summary ............................................................................................. 110

6 Emotional Acquisition Through Behavioral Human Poses .......................... 113
   6.1 State of the Problem of Affect Recognition via Body ......................... 114
   6.2 Data and Methods ............................................................................... 115
      6.2.1 Data Set ....................................................................................... 115
   6.3 Discussion and Results ....................................................................... 116
   6.4 Summary ............................................................................................. 119

7 Emotional Acquisition Through Speech Emotion Recognition .................. 121
   7.1 State of the Problem in Speech Emotion Recognition .......................... 121
   7.2 Data and Methods ............................................................................... 122
   7.3 Discussion and Results ....................................................................... 123
8 Decision-Making Affected by Emotional Feedback Using a Bimodal System

8.1 System Toward the Overall Problem

8.2 Data and Methods

8.3 Overall System Architecture

8.3.1 Bimodal System

8.3.2 Decision-Making System and Self-Regulation of Emotional Feedback

8.4 Results and Discussion

8.4.1 Bimodal System

8.4.2 Decision-Making model Affected by Self-Regulation of Emotional Feedback

8.5 Summary

9 Concluding Remarks

9.1 Concluding Summary

9.2 Future Research

Sammanfattning

References
Abbreviations

\( \text{kNN} \) k-Nearest Neighbors
\( \text{2DMM} \) 2D Method of Moments
\( \text{2DMM-MFCCs} \) 2D Method of Moments of MFCCs
\( \text{AI} \) Artificial Intelligence
\( \text{ANN} \) Artificial Neural Network
\( \text{ARA} \) Adversarial Risk Analysis
\( \text{BN} \) bayesNet
\( \text{C++} \) C Object-Oriented Programming Language
\( \text{CK} \) Cohn-Kanade database
\( \text{eNTERFACE’05} \) Audio-Visual Emotion Database
\( \text{LPCC} \) Linear Prediction Cepstral Coefficients
\( \text{MATLAB} \) Matrix Laboratory
\( \text{MEU} \) Maximum Expected Utility
\( \text{MFCCs} \) Mel-Frequency Cepstral Coefficients
\( \text{MIR} \) Music Information Retrieval
\( \text{MK} \) Matrix-Knowledge
\( \text{ML} \) Maximum Likelihood
\( \text{MLP} \) Multilayer Perceptron
\( \text{NB} \) Naive Bayes
\( \text{RMS} \) Root Mean Square
\( \text{SC} \) Spectral Centroid
\( \text{SCV} \) Spectral Centroid Variability
\( \text{SF} \) Spectral Flux
| **SVM** | Support Vector Machines |
| **VN**  | Vector-Knowledge         |
| **WEKA**| Waikato Environment for Knowledge Analysis |
| **ZCR** | Zero Crossing rate       |
List of Figures

1.1 The Design Science Method ........................................ 43
1.2 Dissertation Roadmap ............................................... 51

2.1 Emotional Machine ................................................... 56

3.1 Basic Agent Loop ..................................................... 80
3.2 Robot actions .......................................................... 82
3.3 User’s actions ........................................................... 83
3.4 Pyramid of objectives. ................................................ 87
3.5 Screenshot of simulator ................................................ 90
3.6 Evolution of $p(M1)$ and $p(M2)$ with the user constantly interacting with the agent. ......................... 92
3.7 Evolution of $p(M1)$ and $p(M2)$ without user-agent interaction. ................................................. 93
3.8 Evolution of $p(M1)$ and $p(M2)$ with an intermittent user-agent interaction. ................................. 94
3.9 Evolution of the expected utility: with interaction (Top), a mixed interaction (Middle), and no interaction (Bottom). .................. 95

4.1 Emotion detection loop ............................................... 98
4.2 Facial Image Treatment ................................................ 99
4.3 Backpropagation Neural Network architecture ........................ 99
4.4 Facial expressions for each individual ................................ 100
4.5 Image data from the eyes and mouth separately .................... 103
4.6 Percentage of successful emotion recognition in mouth zone 104

5.1 Elderly person-agent interaction scenario ............................ 108
5.2 Effects of emotional feedback ......................................... 111

6.1 Social behavior through bodily information .......................... 115
6.2 Sensing of data .......................................................... 116
6.3 Matrix-knowledge: matrix addition of a new image ................ 117
6.4 Accuracy ................................................................. 118
6.5 Recall .......................................................... 118
7.1 Flowchart of the construction of dataset .................. 123
7.2 Comparison of accuracy of different classifiers with all features and with feature selection .................. 125
8.1 Overall human-agent interaction ............................. 129
8.2 Dataset construction of audio-visual information ........ 130
8.3 Overall system architecture ................................. 131
8.4 Agent’s decision-making simulator ............................ 133
8.5 F-measure ..................................................... 134
8.6 Interaction test between the agent and all datasets in order .... 136
8.7 Interaction test between the agent and all datasets randomly .... 137
8.8 Interaction test between the agent and one specific individual .... 138
8.9 Expected utility result of the interaction between the agent and the specific individual ........................ 138
List of Tables

2.1 Confusion Matrix ........................................... 75
3.1 Evolution of being lightened ................................. 86
4.1 Confusion matrix using the complete merged image ....... 102
4.2 Confusion matrix using the isolated zone of eye .......... 104
5.1 Emotions, Actions, and Modified Actions ................... 109
6.1 Confusion matrix ............................................ 117
7.1 Confusion Matrix J48 ........................................ 124
7.2 Comparison of recall for J48, MLP, SVM with degree 2 (SVM-P2) and BN ........................................ 126
Acknowledgements

This may well be my only opportunity to shape this group of words in writing from the bottom of my heart; I might be a bit more loquacious in expressing my thanks than necessary, but it is well worth it because those to whom I am grateful are amazing people.

First and foremost, I would like to thank God Almighty, whose many blessings have led me on the right path, encouraging me and providing illumination along the road of my life.

Thanks to my beloved parents, Marina Rázuri Rodríguez and Javier Francisco Guerrero Valera, for all the time they invested in raising me, time decorated with many happy memories, dreams, and spirit of faith, which has helped me to move forward despite all the obstacles that I have met along the way.

I must offer my profoundest gratitude to Professor David Ríos, as he has inspired me to commit to a career in academia when I had previously been seeing only a blurry future. He extended his helping hand to me, showing me the way of an independent researcher. He shows, day after day, what a bright and untiring scientist can accomplish.

My sincere thanks must also go to Professor Love Ekenberg, a hard worker who is passionate about his research; he has supported me financially through the projects related to my position as a Ph.D. student. He was generous with his time, and many times, he shared his friendly smile accompanied by positive words.

I am entirely indebted to all of my supervisors who helped out with the discussions, suggestions, and advice in this research process. I thank Professor David Sundgren for his relentlessly positive character, infinite patience, guidance, and the faith implanted in me, and for generally being a great person with whom working is easy. I am never going to be able to forget the love of his heart that was revealed in the loveliest Swedish Christmas dinner. I acknowledge Professor Rahim Rahmani for his lofty words that always encouraged me before each complicated presentation, which, at the same time, made me braver and always focused. Additionally, I thank Professor Aron Larsson for his guidance, encouragement, criticism, dynamism, and patience when grappling with a complicated agenda, in addition to the physical distance of separation. They are all friendly, generous, constructive, and good people.

I have had the good fortune of having Professor Ikuo Mizuuchi as my host...
in one of the most valuable research experiences at the Tokyo University of Agriculture and Technology (Japan). The interaction with Ikuo and his group of extraordinary researchers in his lab provided me many successful partnerships, sincere and solid friendship, arduous dedication, and the proof of what a dedicated group of competent professionals can do for the collective benefit of science.

Greatest thanks to Maria Daniela Artigas who has become a special and trusted friend over recent years, we first met when I was working very close to her Lab in a collaborative research project as part of the Ph.D. research in Tokyo University of Agriculture and Technology (Japan).

Special thanks must go to Professor Becky Passonneau for hosting me at the Center for Computational Learning Systems at Columbia University (New York) as a visiting researcher. The experience gathered at the CCLS department was infused with the very essence of academia, an atmosphere of ethical work, boundless curiosity, and unmatched enthusiasm. Thank you to my dearest friends Boyi Xie and Marouan Khata for the opportunity to share with you lovely moments inside and outside the workplace. The scientific and technical explanations of Marouan related to photography will last forever, and I hope one day to be as skilled as he is. The tough times were overshadowed in light of Boyi’s positive quotes in the middle of Grand Central Station (New York); they were able to predict many positive aspects in my future road.

Words can never suffice in expressing how grateful I am to Professor Isis Bonet Cruz who taught me the importance of continuing to fight and the delight of pursuing a dream despite the adversities. She didn’t hesitate for one second in lending me her expertise and intuition to help me solve scientific questionings. There is no way to express how much her stories about continuous fight, pride, and temperance that have occurred in Ciego de Avila (Cuba), her natal town, have meant to me.

I am grateful to Professor Antonio Moran Cardenas for having encouraged me in taking the risk of traveling outside my country to build a better future for myself. I am grateful, again, because I have extended my research expertise throughout the collaborative work carried out with him, and I owe a lot to Professor Antonio for this achievement.

I would like to convey my heartfelt thanks to four special people who have welcomed me and encouraged me since I started to become a part of the department as an active worker: Irma Muñoz for her understanding and her love that encouraged me to work hard and to continue pursuing my Ph.D. imbued with the vision and energy of “Latinos” that we share; Sven Olofsson for his attitude of helpfulness and unconditional friendship; he was always near me with a never-ending smile; Professor Lars Asker, who shared his intellectual knowledge as well as his time in long talks concerned mostly with the cir-
cumstances of life in which the advice “be positive” is the key to success; and Professor Josef Swiatycki, who extended my thinking about my research work in academia through our talks and his company, day after day, in our workplace very late at night. Thank you all; your support was invaluable!

I always keep in mind a phrase that I have usually heard on weekends: “You are killing yourself!”, followed by a sense of tranquility that I have never perceived in others. I can’t think of a more fitting way to sum up the positive encouragement bestowed upon me by Professor Oliver Popov.

I have discovered during my personal journey as a Ph.D. student a staff of amazing researchers and devoted people that as co-workers went on to become very valued friends, and they are and will be part of my heart during all my life. I cannot forget the fun we have had, prolonged scientific discussions, and the most important thing, their friendship, a most valuable treasure that has allowed me to enjoy my work days. My thanks and friendship especially goes to Meshari Alwazae, Khalid Bencherifa, Anton Talantsev, Samuel Bohman, Tobias Fasth, Osama Ibrahim, Angela Westin, Rueben Laryea, Maria Hägglöf, Vamis Xhagjika, Ram Bahadur Gurung, Irvin Homem, Xavier Muianga, Constantino Sotomane, Bin Xiao, Iyad Zikra, Thashmee Karunaratne, Parisa Aasi, Jing Zhao, Hasibur Rahman, Aron Henrikson, Asreen Rostami, Avelino Mondlane, Uvasara Dissanayake, Suliman Ouchen, Orlando Pedro Zacaraias, Jean Claude Byungura, Ana González Marcos, Mohamed El Mekawy, Sameer Bathallath, Naghmeh Aghaee, Isak Karlsson, Henrik Warpefelt, Bernard Bahati, Maria Kalinina, and Workneh Yilma Ayele. The list of friends who provided me kindly help and affection is enormous, and the space available on paper is limited. I apologize on behalf of my memory for any omissions.

I would like to thank Mariela Du Rietz Concha Ferreira who, with a single word, “Hermano”, made me feel like I had a family in Sweden. She has provided me her unconditional support, and her amicable and positive disposition within the academic environment. I am extremely thankful to have Mariela as my friend because her loving patience has also provided me assistance to reach this feat.

Special acknowledgement goes to Serge Thill from the University of Skövde for being the opponent of my pre-defense. He supported me to attend to issues that I might otherwise have neglected through a deeper analysis of my thesis, providing me enlightening comments.

I still have many fond memories about my old workplace at the University Rey Juan Carlos (Spain) where I got to know people with outstanding attributes, providing me their unselphish friendship that I will treasure for the rest of my life. Thanks are extended to Alberto Redondo, Pablo Gómez Esteban, Javier Cano, Javier Gomez, César Alfaro, José María Lavín, Matti Sutinen, Omar Romero Sandoval, Alcides Romualdo Neto Simbo, J. Marie Kiombo,
and Benjamin Osorio. You have my sincere appreciation for your time and collaborative work. You welcomed me into your scientific family; thank you to all. Apologies if I have overlooked anyone, but no worries, the hard copy is still on the bottom of my hearth.

In full gratitude, I would like to acknowledge the Pontifical Catholic University of Peru (PUCP) and the Automation and Control Group for offering me an ideal environment surrounded by collaborative work, in which I felt free to share my research with their members. Special thanks are also given to Gustavo Pérez, Edwin Milton Calderon Mendoza, and Luis Miguel Enciso Salas, who make up an extraordinary group of young researchers and friends at once.

I would like to express my special appreciation and thanks to my home university, Universidad Peruana de Ciencias Aplicadas (Peru), for it is there where I forged my curiosity that is always leading me to discover new paths. My curiosity has been a tremendous mentor, enabling me, day after day, to grow as a research scientist.

My deep appreciation goes out to the members of the examining board, Professor Alessandro Villa (Opponent), Professor Serge Thill, Professor Jelena Zdravkovic, and Professor Jonas Beskow, for their detailed comments.

Furthermore, I would like to thank to everybody in the Department of Computer and Systems Sciences (DSV) at Stockholm University for the warm atmosphere, encouragement, and the commitment to ensuring a great place to work.

"From the standpoint of daily life, however, there is one thing we do know: that we are here for the sake of each other - above all for those upon whose smile and well-being our own happiness depends, and also for the countless unknown souls with whose fate we are connected by a bond of sympathy. Many times a day I realize how much my own outer and inner life is built upon the labors of my fellow men, both living and dead, and how earnestly I must exert myself in order to give in return as much as I have received."

Albert Einstein
Imagine this situation: in ancient Japan, at the starting point of the cha-no-yu (tea ceremony), two samurais are taking part. The environment has a flavor of tranquility, contrasting the qualities of strength and aptitude possessed by the couple of samurais. A simple error within the tea ceremony is analogous to a significant mistake made on the battlefield, but thankfully everything went well with the two samurais in the scene. The situation changes, and a Zashiki Karakuri (a tea-serving doll) will instead take part in the process. The doll does not have any expression; it is an inanimate mechanical automaton. It simply carries a teacup in a tea tray to the guest, and when the guest stops drinking, the Zashiki Karakuri turns around and returns to the host. The process is very simple; through their motions due to the mechanics, the doll acquires something similar to humanity understandable by the samurais. But, what would happen if the doll chooses the wrong way, drops the tea cup on the floor, gives the cup back to the guest, etc. Impolite actions could provoke emotional behavior between the samurai and translate the lived experience into an uncomfortable situation. The Zashiki Karakuri is only an inanimate automaton without any behavior according to social norms, and has a null individual/automaton emotional connection pattern. On the other hand, what would have happened if the doll made decisions based on the subtle expressions and feelings provided by the samurai (voice, facial gestures, or behavioral expressions), within the social interaction elements that are present in real life. This, of course, would have been impossible for that era. Maybe the connection between the two individuals had not diminished, and their rejection of the doll had not occurred. Two individuals, an automaton, an environment, and emotional signs not recovered as a result of a few misunderstandings in the interrelation social loop are just one concrete example of the power of sharing emotions with machines.

We are currently working on the foundations of what will be a significant challenge in the near future, the understanding of human emotions by machines. We are in the first stage of a new age based on multisensory connectivity, in which all of the wearable and smart devices can share information collected by their sensors. Robots, agents, entities, or whatever we want to call them, are also part of the new wave of multisensory connectivity, in which the critical information provided by humans is shared by machines with a rapid pace, en-
abling decision-making according to the collected information. But we could also consider that the great percentage of recovered information from humans might be handled as inputs with emotional connotation, and from here on out, manage the emotional information to make the evolving relationship between humans and machines closer.

Finding the connection bridge between emotions and what machines can do with them at a desirable level could endow them with skills with a certain degree of humanity, helping humans who are sharing the same emotional connection. The study of emotions in machines has become important from its entry at the doors of the millennium. This was the break point in time wherein the artificial intelligence (AI) and human-computer interaction fields started to be broadly researched, focusing on the new technologies that advance our basic understanding of emotion and its role in human experience. In AI machines, the distinction between emotional and affective content is linked at a global process of the recognition and conception of an emotion as part of a whole within the affective process. This special aspect allows us to build more realistic engineering designs in agents, which will be equipped with computational models of affective development [1]. Emotional interactions between users and machines within intelligent environments serve as fertile ground for further studies on affective interaction, which could open up a promising path related to the machine’s understanding of complex human emotional behavior [2], [3], [4]. To create an inanimate entity that may develop behaviors at least somewhat similar to those of human beings, it is reasonable to believe in an architecture as a role model in which each part forms an essential part of a unified whole. To think solely in terms of emotional architecture with a level of understanding of the human cognitive system entails the recreation of mechanisms similar to human minds. More precisely, architecture is the fundamental organization of a system embodied in its components, their relationships to each other, and to the environment, and the principles guiding its design and evolution [5]. Human emotional-centered systems as carriers of emotional architectures are a matter of study in diverse scientific fields such as neurology [6], medicine [7] and psychology [8], to name just a few. They are all aimed at seeking life-like or human-like interaction to enhance the human interactions in the physical world [9], [10], [11], [12], [13], [14].

Machines can perform accurate and quick calculations with rational, rigorous, and logical behavior, but at an emotional level, machines would have to achieve one or more pillars of the “affective loop” to recognize emotion, understand emotion in context, and express emotion naturally [15], [16], [17], [18]. They could use data from emotional cues with some autonomy to maintain the
social connections with human beings, building the foundations of what could be called “emotional feedback”. Machines do not actually feel emotions yet, but they can appear as though they do. Assuming that machines could somehow accrue some emotional skills, they could manage the control of criteria in their actions and through judgments of value, adapt their behavior to improved situations and pursue support in a selfless and helpful manner, with this process being a bridge for the delivery of optimal support to humans.

The future emotional skills in systems will create a new dimension in human lives and new ways of interacting with technology. Current research is just beginning to close the emotional bonds in machine-human relationships, which is already becoming a promising reality. The influences of external variables provided by an emotion perception engine could serve as an effective pathway to regulate internal conditions within decisional processes in a machine; this capacity would allow machines to be prone to help people within an emotional connection. Emotional systems have already gained some level of interaction with humans within a wide range of application scenarios such as self-driving vehicles [19], butlers [20], nursing care [21], elder care [22], assistance (pets and people) [23], and personal entertainment and companionship [24], but so far it has been a difficult task to understand the human state of feelings. For the time being, intelligent systems are being built to provide information and enable humans to make better decisions, but without taking account of the most basic human needs. To be effective, entities such as machines, software agents, etc., must address human perception and the emotional nature of our response, and proceed accordingly.

1.1 Motivation

Human-machine interaction has made significant progress, but machines are still some way from being able to interact with humans in a seamless manner [25]. However, it is interesting to consider the understanding of feelings in machines and the use of these feelings. It is important to tackle the question about the concept of emotion and think about the “feelings” in machines as in humans, as a key to imparting value, intuition, and deep meaning to emotional machines. The understanding of the complex emotional realm of humans and its physical indicators have been carefully studied and analyzed in order to understand and track emotional experiences within social behavior in order to obtain a level of understanding, enabling us to employ this emotional knowledge in the design of artificial entities [26], [27], [28], [29], [30]. The human sensory system is also accompanied by various emotional channels, in which combinations of data from different senses can facilitate emotional prediction;
it is merely pure information from human behavior that an agent can easily handle with a high level of precision. The data encoded from emotional situations might provide to an agent different ways to discern information about human processes [31], [32], [33], [34]. There is an extensive research literature and a number of well-established tools for measuring non-verbal communication. Most of the decryption of non-verbal signals aims to assess accuracy in the recognition of emotions as expressed by others and is commonly recognized in humans as a decoding skill, focusing on facial gestures [35], [36], voice [37], [38], [39], kinetics of movements and postures [40], [41], [42], [43], and the fusion of all of these factors within emotion analysis systems [44], [45], [46]. Multi-sensory information in humans is a rich source of data in order to decrypt affective states [47], [48].

As an example of emotional systems in the foreseeable future [49], machines will be more present in our daily lives, helping us at work and at home as personal assistants; they could make lives more convenient. They will be collaborating with people and other intelligent systems, and they will coexist with a variety of entities. Through computational power, machines could achieve realistic-looking depictions of human actions, emotions, and even decisional processes that may assist them in reacting emotionally in day-to-day situations, recognizing the key elements around human emotional life, and so forth. Maybe the future may bring with it an even more realistic experience with synthetic companions who could mitigate the demands of living in an overpopulated world.

Recently, the field of cognitive processes has shown that emotions may have a direct impact on decision-making processes [50], particularly in areas such as affective decision making, neuroeconomics and affective computing. These disciplines are based on this principle and are impacting the general area of affective robotics [51]. Emotions are equally important in human life as the most reliable indicators of socialization. They determine how we think, how we behave, and how we communicate with others. The mixing of these concepts gives rise to what is known as a “social agent”. Within a group of computationally intelligent agents, a social agent is capable of making decisions influenced by emotions and interacting with humans. Based on the fast evolution of research and innovation of new technology, it is reasonable to imagine a probable future in which machines exceed human performance, with enhanced capabilities [52] at some level of emotional understanding and decisional capabilities [53].

Emphasizing the performance of a “social agent”, the prospect must be op-
erated autonomously in a real environment through the context understanding in which it evolves the anticipation and adaptation to changes, autonomous behavior in decision-making (individual and group), as well as reciprocal human-agent relations, giving priority to the connection with emotions. “Particularly for mobile robots, we need to have something similar to emotions in order to know, at every moment and at least, what to do next” [54]. To achieve an interaction that is human-like, the combination of a self-regulated emotional system and decisional system could be mandatory.

The vision towards the maximization of human happiness is strongly supported through the building of links with machines, and certainly this idea has not yet been translated into a form sufficiently rigorous for the Artificial Intelligence community. The road of self-supported emotional interaction with humans could enable a machine to adjust its behavior, maximizing its goals adeptly by taking into account the information collected around the changing environment. Along the way of this interaction, the machine would truly be perceived as an intelligent being to which humans can relate, at least in some respects within the same environment, evolving with them. A machine able to express, recognize, and communicate its own feelings and those of others is capable of enhancing human-computer interaction and aiding related research in surprising ways [2].

To provide machines with emotional decryption and crack the code of the human emotional dimension, it is necessary to learn about the human processes to convey emotional messages provided by multiple channels, e.g., intonations in the speaker’s speech can be affected by the dynamics of face expressions; these sources are combined in a complex thought process to obtain a particular emotional response. Expressions, gestures, vocal inflection, posture, and gait all say something to those around them, often in stark contradiction to the spoken words, e.g., social psychology has shown that conveying messages in meaningful conversations can be dominated by facial expressions, and not by spoken words [55]. The case of emotional cues as found through facial expression has also shown connections in a number of cognitive tasks, e.g., in a noisy environment, the lips can contribute greatly to speech comprehension or a simple action of knitting one’s brows makes clear our disagreement about something. Is clear that more than one emotional source can bring more precise emotional information (embodies some level of classification) in order to lead to an intelligent action. Focusing on speech, a rule to follow in emotional decryption is not yet established, but emotional cues in speech are mandatory to learn about some human emotions reflected. It is known that emotions cause mental and physiological changes that are also reflected in uttered speech [56].

35
As humans emanate multiple emotional signals arising from multiple emotional channels, machines must also have the capability of processing multiple streams of information and deciding which could be better detected, with more emotional sources and different modalities to analyze the complementary information, since the performance of emotional decryption increases when various emotional cues can be complemented.

A machine capable of developing empathic behavior and acquiring information through social interactions to seamlessly connect in an emotional way, taking part in the elder care environment, is usually called a “companion”. This type of machine will be able to operate pro-actively and autonomously adapt its operations to fit changing requirements and evolving users’ needs as well as emotional states, via learning during interaction. The data decrypted from emotional states is useful to produce agents that fulfill the emotional needs of seniors; the connection with real emotional indicators is a crucial element to select what kind of responses and actions need to be taken related to aging for home assistive technology. A companion is capable to help, and it could generate a smart space based on all of its sensors, integrators, and processing capabilities integrated and capable to capture information, e.g., a machine with these qualities will provide considerable support to the independent life of the elderly. The concept of companion technology underlines different important aspects such as empathy, emotion, social intelligence, and the related technology for independent living well into old age. Machines within a companion environment that fulfill their function as caretakers will learn to adapt to the moods and actions of their patients. Over the course of time, they may also develop stronger bonds, according to relevant chosen behaviors that are connected with different aspects of daily life such as empathy, emotion, social intelligence, etc., enhancing all information sharing, which is the primary driving force. The companion must show dynamic and autonomous behaviors, increasing the user’s feeling of interacting with an intelligent, emotionally responsive personal entity. In truth, to shorten the elder-agent connection link, all efforts must be focused to generate “engagement” in order to increase its willingness to be more natural and engage in acceptable interactions. This may be done through understanding the message from a broad range of emotional data, probably both intentional and non-intentional, that are delivered in a non-verbal communication. Indeed, developments in emotional and decisional models may be seen as a way to experiment with new alternatives of social interrelations. A proposal for a behavioral model in a companion could be endowed with the capability to use emotions in decision-making as a link in a chain of emotional reason-
ing, permitting them to easily socialize with humans. The actions from the individuals could be registered and processed in a probabilistic way in order to predict the individual’s subsequent actions. As a result, this simple model produces intelligent decisions that allow the machine to analyze the consequences of its chosen actions based upon the external emotional status provided by humans. The machine not only takes into account signals from its current state, but also considers the consequences of its future actions while simultaneously predicting the user’s response affected by a closed loop of emotional feedback provided by humans in a unique environment. The accompanying emotional functionalities could be implemented using the development of new methods for the analysis of decision-making, affecting all of them with human emotional inputs. An emotional companion will be able to interact in a natural manner with human beings, providing societal interaction affected by emotions. Thus, individuals who reach an elderly age could require the support of a companion; this is a quite strong reason to create new models for fostering the self-support, cognitive pleasure, and self-fulfillment of these people. Human beings and machines would be attached by a strong sense of emotional factors up to the point that, somewhere in the not-too-distant future, we will be living in a hybrid society; a society where humans and machines will be inextricably interwoven.

1.2 Problem Statement

Aging is related to the fundamental biology of age changes in which physiological functions in various organs and tissues lead to an increased probability of death. During this period, sometimes humans cannot manage by themselves and require additional care. The world’s population is aging, and policymakers in different countries have broadly agreed about what has to be done to better promote improvements in eldercare. The 2015 Ageing Report [60] shows results of rates of population pyramids related to age structure of the European Union (EU) population. The results predict that the population will change dramatically until 2060, and that elderly people will account for more than any other population. This is because compared with the populations of children and middle-aged adults, people aged 80 will become more numerous. This trend is further influenced by the fertility rates that will also start to fall. Europe is currently the oldest continent, with the highest old-age dependency ratio, and will remain so in 2060. Other zones of the world are increasing the aging of their populations in a dramatic way, in which the old-age dependency needs could climb to high levels. To be able to give elderly people the benefit of living longer, the fulfillment of their special needs and requirements must be taken into account. Against that backdrop, a technology shift brings with it
incredible possibilities, but also a new set of challenges. Currently, around the world it is still possible to envision a sufficient number of human caregivers for the elderly, but this reality could change in the future due to many factors, including the declines in fertility (fewer children entering the population and people living longer), rapidly aging populations, demographic changes, and the low subsidies for elder care that discourage people to find caregiving to be a fulfilling and long-term profession, just to name a few. In view of the increasing demand for care personnel, societies around the world have to find strategies for addressing these challenges.

According to a new study by Grand View Research 2014 [61], the market for personal robotics will experience an accelerated growth that will bring high revenues near the year 2020. Military and defense applications constitute a significant part of the market, with the mobile robotics market for service applications, wherein the growth will be faster within the forecast period. In the ABI Research report 2013 [62], it was reported that the market for personal robotics will reach a turnover of more than 19 million USD annually by year 2017. With the era of personal robots just beginning, it is likely that the arrival of the 21st century could bring an increased number of roboticists around the world, in which the arena of emotional caregiving is one in which the future of robotics and their architectures with special caregiver requirements could blossom. A non-human entity that could serve as a caregiver is supported by the fast-developing technology market that is increasingly driven by the introduction of novel and improved products. Many of these products involve the mastery of software and hardware complexity that could be pivotal towards exploiting the caregivers’ capabilities, as social and emotional factors are key components for designing believable agents. A significant percentage of households across the world are not exempted from the aware attention of a caregiver, centered in one or more family members. Despite the level of caregiving (physical, cognitive, mental, sensory, emotional, developmental, or some combination of these), many countries find themselves struggling to provide the funds and staff necessary to supply care for the majority of these cases.

A possible solution of the elderly care problem in the next generations could be linked with the extensive research into emotional assistive technology for the aging, in which the design of agents with a certain autonomy in self-care and with decision-making capability will be mandatory. Connected with information about the user’s feelings, this could provide us a wider vision of the future personal emotional well-being systems. The needs of companions to play a key role in caregiving for elders has seen an exponential growth [63], as shown by the concern of the countries who realize this and are investing
significant funds into the development of such systems. Perhaps the emotional systems could play a substantial role in caring, to able to talk to us, entertain us, and respond sympathetically to our emotions and from this point manage decisions, such that we will never need to be lonely, as long as we have our own care system support.

1.3 Research Question

The central core of the research in this dissertation was guided by the following research question:

Would an architecture composed of emotional feedback provided by humans and an internal decision-making engine lead to an agent generating a group of ‘Decisions with emotional connotation’ that may make the link established between the agent and the human closer?

This research question was answered through a group of requirements related to establishing, developing, and understanding with the support of isolated simulations of each subpart towards the assembled collection of the entire emotional behavioral model. The entire model should be capable of delivering a timely and reasonable response, especially given the emotional features provided by humans within an emotional feedback framework and with an acceptable degree of autonomy in their interactions, based on decision-making models that employ an evolving forecasting model. The requirements are realized against an architecture that demonstrates the feasibility of building a system that embeds decryption of human emotion and decisional capacities, allowing a better communication between the human and the companion system (robot, software agent, etc.). This vision is able to support the bases of a future with symbiotic relations between humans and machines, which will aim at the design of new interactive technologies based on new theories and models of human cognition and emotion, non-rational decision-making, social behavior, and spatial and temporal perception and processing.

1.4 Research Methodology

From this part of the chapter onwards, the content is devoted to the analysis of the philosophical foundations of the research covered in this dissertation. The dissertation’s position is analyzed, taking into account the paradigmatic nature of the scientific enquiry process. The chapter shows further how the scientific enquiry process informs about the structure and outcomes of the research
contributions of which it is composed. In common parlance, “research” is the quest for knowledge by humans, in which enquiry actions are performed to search for new evidence with the application of scientific procedures. Sometimes this does not imply reinvent the wheel; to make the concept more specific, it could be dealing with a problem that other researchers have wrestled with before us, giving rise to new discoveries or their improvement. The systematic and scientific approach employed in research can describe, find the meanings, make forecasts, and control an observed phenomenon, involving inductive and deductive methods [64]. Most of the scientific enquiry actions are centered on solving the known unknowns, giving way to the research questions that are useful and relevant during the entire course until the termination of research. The scientific enquiry is not carried out in an idealized environment, but all the time the research actions will face specific and changing contexts [65]. To work with known unknowns over the structures of scientific paradigms could involve an action of making efficient selection or design decisions. All alternative options are progressively narrowed down in order to accomplish a task until its understanding has been achieved [66]. Scientific enquiry is generally described as a series of collections of methods with slightly different steps; all of them aim to answer the unknown. The research methodology itself is the underpinning of systematic approaches to provide a solution for a research problem [67]. A research paradigm focusing on the research community is based on a set of shared assumptions, values, concepts, and practices [68], where all of them are related to the ontological, epistemological, and methodological aspects that compose a scientific enquiry. The essence of the paradigm could be the process of thought that the researcher has about the development of knowledge. Through this, the scientific enquiry is molded in the field and defines the methods and the applications of each of them [69].

1.4.1 Philosophical Assumptions

Scientific inquiry is directed at the construction of knowledge through several philosophical assumptions and in the context of a scientific community. Ontology was born in philosophy; it is related to the systematic definition and study of the nature of existence, or being [70]. The assumptions of being or reality, what currently exists, and how it should be considered and interpreted give way to the ontological dimension of research. Reality will be treated as a form of knowledge that develops during the research process [71], [72], defining the epistemological aspect of the enquiry. The collections of methods are focused on the acquisition of the knowledge and the information of the outcomes, thus forming the methodological aspects. The collections of applied methods cho-
sen by the researchers are shared by the community grouped into philosophical assumptions as a result of the course of the enquiry [72]. The philosophical assumptions are the paradigms used by researchers to use to gather, analyze, and interpret data within a research investigation. As a result, the definition of a clearly philosophical position that could explain the basis and conduct of the enquiry to the broad community is mandatory [73]. The philosophical assumptions underlying the fields of computer science and information systems come mainly from interpretivism and positivism [69].

**Positivism**

The *positivism* paradigm of exploring social reality is based on the conjunction of observation and reason which make possible the understanding of human behavior. The experience of senses brings the true knowledge constructed from observation and experiment. Positivism, from the ontological point of view, takes into account that reality is capable of being measurable, using properties independently of the researcher and the instruments they employ; it is remarkable that knowledge has the characteristics of quantifiability and objectivity [74]. Positivism aims to explain and discover general laws that arise from relations between observed phenomena, focusing on the cause and effect. The results of the positivist explanations must be void of context and generic throughout the entire research community.

**Interpretivism**

Interpretivism takes an approach based of the reality fabricated over the people’s subjective experiences from the external world. It could adopt an intersubjective epistemology and the ontological belief that reality is assembled by social connotation. Interpretivism is guided by the interpretation and observation of events. To the extent that observation collects information, the interpretation acquires meaning of the information recovered, making inferences or value judgments between the information and some abstract pattern [75]. Interpretivists state that reality is multiple and relative from the point of view of ontology and epistemology [76]. The multiple realities are linked with a multiple group of meanings; this assumption makes more difficult the interpretation of terms of fixed realities [77]. Ontological perspective is not without human influence, and the cause-effect is not able to be carried out without the social factor, as that could impact the outcomes produced. The epistemology perspective takes into account knowledge as being subjective, a result of observation with less contemplation of the researchers involved in the process. Interpretivism does not have rigid structures; the structure is more personal.
and flexible [78], predisposed to collect meaning through interaction with humans [79]. The research and researchers show interactive and interdependent behaviors, maintaining open knowledge throughout the study with the support of informants. Interpretivism has assumed the collaborative approach with beliefs that humans have adaptation skills, and nobody can acquire previous knowledge of time and context constrained to social realities [76]. Interpretivism seeks the understanding and the interpretation of the several meanings of human behavior instead of the generalization and predictions of causes and effects [80], [76].

- **Choice of Paradigm**

  The contributions resulting from the research carried out were focused on the realization of a set of simulations, which are mathematical models and algorithms, in order to prove the viability of the proposed system. As the science of problem solving in this dissertation is within the field of computer science, all of the contributions have the philosophically positivist point of view. The level of improvements and performance were measured through simulations within a design verification and testing framework, giving us the truth or falsehood about the existing hypothesis or the contributions grouped in this dissertation. The works that build this dissertation are based primarily on a close examination of the interpretive paradigms of social computing, taking into account that all of our own contributions are treated with objectivism, objectivity, and the proof of theories related to abstruse or concrete knowledge. The contribution of this dissertation acquires a positivist flavor to the extent that the experiments are focused on to verify the results obtained. The positivist aspect begins from a hypothesis to be tested, which is linked to the question: “Could an external human emotion that is achieved change the behavior in an agent’s decision?” On the basis of this logic, the research is a sum of tests or set of indicators that try to prove the truth of this belief by measuring the target outcomes, in this case a group of decisions affected by emotions. The collected data from the test stages with simulations tried to draw knowledge about the positive influence of emotional decisions within the interaction in a human-machine loop. The use of simulations in this context suggests the need for a simulator to prove adjuncts to manage and evaluate, analyzing the full functionality of the entire architecture. The results achieved are not based on only mathematical models, machine learning algorithms, set of rules, or psychological theories that have been represented through software tools. The results provided by the research are generalized to the single objec-
tive reality. All of the isolated simulations have sought to deal with the fact that the model can be divided by parts and that each part can deal with non-specific scenarios.

1.4.2 The Design Science Paradigm

Design science is the scientific road to pursue the creation of an artifact. The artifacts can be used by the people to solve and understand practical problems concerning all the community and that provoke a general interest. The research conducted under the design science paradigm can be handled by more than one scientific method, taking into account the step that is being developed and the goal to be attained [81]. The artifact as an object is made by humans with the purpose of addressing a practical problem [69]. The evaluation of a designed artifact highlights its ability in the solution of a practical problem in accordance with the stated requirements of the artifact [82]. The classification of artifacts is varied and could be based on their function and use or the type of knowledge they express (explicit or embedded). Computer Science and Information Systems have a great quantity of artifacts, extending from algorithms, logic programs, formal systems over software architectures, information models, design guidelines, prototypes, and production systems.

![Figure 1.1: The Design Science Method](image)

Fig. 1.1 shows the method followed in design science to lead an enquiry process. The method includes five main activities extending from problem investigation and requirements definition, through artifact design and development, and finally, to demonstration and evaluation. This dissertation is based on a list of contributions that follow the structured scheme of activities that constitutes the design science process.

43
1.4.3 Design Science and the Scientific Enquiry

Using the methodological structure of design science, we assembled a group of methods applied in each stage of the process in order to achieve the artifact. The artifact in this dissertation is a system that will track a human’s affective state based on emotional indicators routinely found in human-human interaction with the purpose of modifying the forecast actions of a decision-making model based on adversarial risk analysis (ARA). Through learning of the individuals’ emotional patterns, and information about the surrounding environment, agent’s actions, and individual’s actions, it is possible to generate a continuous flow of agent’s decisions that are pertinent to the individual within an emotional feedback context.

Problem Explication

The problem addressed in this dissertation related to a system that will support an emotional companion capable of making pertinent decisions influenced by the external emotional status provided by humans is outlined in Chapter 1 and is detailed in the group of publications shown in Section ix. The literature survey about current contributions towards affective computing architectures that support decisions affected by external emotions was undertaken with peer-reviewed contributions selected that documented approaches, from testing research hypotheses to testing the validity of claims and results. This point of view constitutes the establishment for the bases of future solutions that could embrace both humans and machines as two resources within a single interrelation and collaborative environment, taking into account emotional information.

Artifact Requirements

The artifact presented is comprised of a group of mathematical models, algorithms, sensing and simulation techniques, and targeted approaches for moving the current state of affective computing research towards an emotion-decisional system that could bring support for humans through a machine. The requirements for each contribution are specified as a result of the examination about the existing state of research and current related works in affective computing. These are detailed and discussed in Chapter 2 providing a general overview of all of the requirements related to each area in the research conducted. The requirements are developed from Chapter 3 to Chapter 8 and, as a whole, culminate in the establishment of the final artifact delivered in this dissertation.
**Design and Development of the Artifact**

The artifact was developed in several stages, taking into account that it is a union of two models (decisional and emotional); each stage as a contribution is a submodel within the entire model. The experimental framework used in each stage provides us information about what were the improvements applied to the overall artifact. The groups of contributions in each publication are the pieces that build the artifact; all of them are detailed with the corresponding design considerations. Each individual content is provided in each chapter, starting with Chapter 3 until Chapter 8, showing the implementation-based approaches with the details of each specific implementation.

**Demonstration of Artifact**

The demonstration aims to show the future use of the artifact depicted in a real-life case, providing the feasibility of the artifact. The experimental framework is proposed in each contribution to support the theories involved in the dissertation to build the final artifact and test its validity. The experimental framework includes the experimental evaluation over each component that tests the approaches, the components, and the applicability of the artifact to conduct scientific enquiry capable of facing real-world problems. The entire architecture is composed of two models that are fused to generate in an agent a decision influenced by an external emotion. The stage related to the model based on Adversarial Risk Analysis (ARA) is depicted using the data from a simulator that represent the behavior of an agent that is capable of inferring the current state of its environment and what its user is doing, through a system of sensors. The data obtained are probabilistic and depicted in evolving behavior graphs. Several methods of experimentation and evaluation in automatic emotion recognition are focused on the data provided from facial expressions, speech, and human body poses, in order to create a vector of features with the aim to analyze the vector in a bimodal system composed of two baseline unimodal classifiers and a multi-classifier. To evaluate the automatic emotion recognition stage, a group of measures for machine learning have been used to establish a rigorous comparison among the methods and algorithms developed, in terms of obtaining successful individual classifiers used in the final model. The final fusion of both models is based in a group of self-regulation rules between actions and emotions named “Emotional Feedback”. The outputs are decisions with emotional connotations achieved by processing and fusing data from the two models with a fusion program that selects the agent’s action related to the emotional information of the user and that is depicted in a group of graphs of the evolving behavior of the agent and user.
Evaluating the Artifact

The artifact evaluation determines how well the artifact solves the problem by taking into consideration solution objectives (i.e. the definition of all requirements). The evaluation strategy used in this dissertation was to carry out simulations within platforms and simulation environments from approaches related to some parts of mathematical models from decisional stages, stages of emotional acquisition, and the fusion of the two stages within the concept of emotional feedback. These simulations and testing experiences were developed using object-oriented extensible programming languages including C++ and Python, the numerical computing environment MATLAB, and the open-source environments in data mining and machine learning provided from the Waikato Environment for Knowledge Analysis (WEKA). The expected result was focused on the convergence of a global approach through the testing of the different developed stages from the entire artifact. The aim of the dissertation is to build foundations for decisional systems with emotional behavior that could serve as support for robotics to care for the elderly, through isolated simulations in several parts of the entire model to evaluate the connection between them. Each chapter of the dissertation shows a structural evaluation approach in each contribution. All of the results are discussed and evaluated taking into account the relevant approaches.

1.4.4 Contributions

The author’s contributions are listed in Section xi. The overall contribution aims at combining and expanding recent developments in affective computing, emphasizing the combination between decision-making and emotion acquisition models. We will consider agents that could serve as emotional companions, which may demonstrate a more or less social behavior, within the unifying model to make decisions supported by the tracking of emotional state in humans. The main contributions are summarized as follows:

- Decision-Making Model of an Agent

An autonomous agent needs to have some special capabilities that give meaning to the term “autonomous”, such as safe behavior, robustness, perception, sensing, and decision-making. These capabilities go hand in hand with concepts and technologies that are necessary to take into account in order to design them. This first part of the research has been focused on decision-making as the important issue towards the autonomy level in an agent. The model was described in Chapter 3 and in publications [1], [2] and [3]. Basically, the context of the problem is viewed as a decision analytic one, in which the principles and procedures that
employ the adversarial structure, and other information available, are
used to assess the probabilities of the opponents’ actions. Assuming
that the autonomous agent imperfectly perceives its environment and
the actions performed by an individual, the model is available to sup-
port the decision-making process carried out in the interaction. The ap-
proach has a decision analytic flavor based on adversarial risk analysis
(ARA), but includes models forecasting the individual’s behavior and
its impact on the surrounding environment. These forecasting models
are then mixed. They are combined with the utility function to approxi-
mate the expected utilities of various actions. The alternative is chosen
randomly with probabilities proportional to the expected utilities. The
dissertation demonstrates that this model can be combined with another
model that can track the feelings of the same individual from whom
the forecasting behavior is taken, as shown in publications V and VIII
and Chapters 5 and 8. The results showed that the agent is capable of
improving its social interaction through the selection of more pertinent
actions by changing its behavior, according to the affective state of the
opponent.

- **Emotional Acquisition Through Facial Gestures**
  In order to be able to influence humans correctly, the autonomous agent
  needs to obtain reliable information about their emotional states that
  would allow it to react accordingly. This dissertation takes as prepon-
derent the emotional connotation shared within machine-human inter-
action, that is why it is important to take advantage of as many emotional
indicators from humans as possible in order to make machine-human
communication richer and more effective; for this reason, a group of ap-
proaches to automated identification of emotions are developed. Chap-
ter 4 and publication IV of this dissertation have addressed the issues
of how to provide the emotional perception, focusing on the individ-
ual’s emotional gestures by a method to detect six basic facial emotional
expressions by using an artificial neural network (ANN) within merged
images. Chapters 5 and 8 and publications V, VI, and VIII used the same
approaches, demonstrating through the testing stages the feasibility for
integration with the agent’s decisional model.

- **Decision-Making Model Affected by Emotional Feedback**
  With the data provided by the opponent’s face and the decision-making
  model of the agent, it is possible to generate forecasting of the agent’s
decisions with emotional connotation as a kind of emotional decision,
using the bridge of emotional feedback introduced in Chapter 5 and
The agent’s emotional decision is adapted to emotional information from the individual that is faced. In this case that the evolution of the agent’s behavior toward the individual’s behavior takes a positive path, taking into account that the reactive response between the emotion and action is dependent on the positive or negative influence that is shown, and will be the trigger of the corrective emotional output assumed by the agent; the positive behavior can be seen as a sort of a self-regulation between actions and emotions. The behavioral model for an autonomous agent based on “emotional feedback” is tested with some modifications, a fusion of two emotional sources as face gestures, and the voice from the individual in a bimodal system. The evolution of the agent’s behavior is proved using rule-based emotional feedback with self-regulation as shown in Chapter 8 and publication VIII.

- Emotional Acquisition Through Behavioral Human Poses
  Chapter 6 and publication VI have addressed the decryption of emotional information through human body poses, another source of emotional knowledge that the decisional model can use. This dissertation shows the possibility of managing emotional information by various cues of emotion provided by humans. The data used are collected from videos of human poses for the purpose of studying human reactions to emotional body language. This is challenging to do because of the large variations in motion activities under all kinds of emotional contexts. From different viewing perspectives, the changing human poses can provide an inestimable source of emotional cues. To overcome these challenges, the method applied for the analysis of emotional behavior is based on direct classification from the sum of pixels in two-dimensional images of body poses that were previously processed. To reach an accurate classification, several machine learning algorithms are tested to select the best one according to the accuracy achieved. Part of the publication VI is additionally an extension from Chapter 4 and publication IV.

- Emotional Acquisition Through Speech Emotion Recognition
  A method for recognition of emotions based on speech analysis can have interesting applications in human-machine interaction, taking into account other sources of emotional information to feed the decision model, as a sense of emotional vision of an agent that interacts with an individual. A methodology to recognize six basic universal emotions based on speech characteristics in time and frequency of emotive speech is described in Chapter 7 and publication VII. The method leads to a pa-
rameterization of audio data for the purpose of automatic recognition of emotions in speech generation and analysis features from Music Information Retrieval (MIR), using the MATLAB program. From the experimental results, it was possible to recognize the basic emotions in most of the cases. The method allowed the selection of the sort of features that contain a high level of emotional information, allowing the removal of redundant and noisy features. Preprocessing and feature extraction methods are used in Chapter 8 and publication VIII focusing on emotional cues from an audio-visual scheme. These methods are applied in a bimodal system fuelled by emotional cues provided by the facial gestures and speech from an individual based in fusion techniques.

- **Decision-Making Affected by Emotional Feedback Using a Bimodal System**
  
  According to the main idea, an emotional companion will be able to interact in a natural manner with human beings, providing a social interaction affected by emotions. With this, it seeks to create an emotional attachment that could be handled by machines, which required creating new models for stimulating cognitive pleasure and self-fulfillment for people who depend upon a companion. Chapter 8 and publication VIII describe the development of the entire system that can track a human’s affective state using facial expressions and speech signals to modify the forecasted actions of the autonomous agent or companion. The dissertation shows that the entire model should be capable of responding given the emotional features provided by humans within the emotional feedback framework. The system can show the possibility to act independently even with incomplete knowledge, based on decision-making models that use an evolving forecasting model. To support this, a method for a union of three classifiers in order to decrypt the emotional cues from audio-visual data is introduced through the form of a bimodal scheme of classification. The emotional output information obtained from the ensemble system feeds the decision-making model of the agent, imparting to the agent the power to make a decision included in the group of affective decisions. At this time, all of the models work in unison to establish the emotional decision support model.

1.4.5 **Dissertation Structure**

This dissertation, in addressing the issues surrounding the decisional-emotional support system for a synthetic agent, continues with a discussion surrounding related work and the current state of research in Chapter 2. This identifies
the critical shortcomings of existing research and establishes a set of requirements that are necessary to be met in order to move toward the decision support system affected by emotional connotation. Chapter 1.4 discusses the methodological approach to the research contained in this dissertation along with the encompassing philosophical assumptions. These are discussed to the extent of their contribution to the conduct of the research and the subsequent implications for the results obtained. Chapter 3 presents the decision-making model that infers the user’s actions and environment’s states. The model is an important component within the entire architecture because will provide a continuous flow of decisions, which are directly affected by the external emotions. Chapter 4 goes on to describe the process of emotion recognition from facial gestures through several testing stages. The data provided by this stage serve to modify the agent’s decisional model. Chapter 5 will explain the behavioral model for an autonomous agent capable of improving its social interaction by changing its behavior according to the affective state of the opponent. We address the problem with the proposal of combining two models: an emotional model and a decision-making model for an agent toward the emotional decision. Chapter 6 addresses the problem of emotion recognition through human body poses, presenting an approach that can predict six basic universal emotions collected by responses linked to human body poses, from a computational perspective. The techniques developed in this chapter will improve the approach developed in Chapter 8. Chapter 7 develops an approach that is supported by machine learning techniques to identify six basic universal emotions from non-verbal features of human speech. The analysis is carried out in order to collect external emotional information from individuals facing an agent, toward the usage of emotional cues in the comprehensive decision-making model. Chapter 8 shows the function of the entire system through isolated simulations. The system uses a fusion of the data from two baseline unimodal classifiers built on an emotional multi-classifier to decrypt human emotions from two different sources. The outputs directly impact the model that supports the decision-making process of the decision agent. Finally, Chapter 9 discusses the conclusions and points to future work moving forward from the contributions in this dissertation. An overview of the contributions contained in this dissertation is shown in Fig. 1.2.

1.5 Summary

This chapter has provided a general introduction to the research included in this dissertation. Firstly, it has addressed the main motivations and considerations about the need for an architecture with emotion recognition and capabilities of decision-making as components of intelligent behavior that shape the research
Figure 1.2: Dissertation Roadmap
Next, the research aim was identified based on the established definition of the research problem and motivations, in which the discussions have centered on the design of agents with emotional and decisional capabilities to support the care of the elderly. Subsequently, this chapter has provided a research methodology related to the design science that shows the artifact and the research process. This chapter also discussed the point of view of philosophical perspectives of social computing, and the rationale for selecting design science-based research is provided. The design science paradigm is discussed, taking into account the method followed in design science to lead an enquiry process that shows its nature, principles, philosophies, and types of outputs. Finally, this chapter concludes with the presentation of the main contributions and the structure of the dissertation to be followed.
2. Related Work

We know about the attraction toward technological advancements that could help machines to perceive, interpret, express, and respond to emotional information. Also appealing is the prospect of their embedded unique personality, knowledge, and experience based on Artificial Intelligence (AI) algorithms to learn and adapt to each circumstance, providing an interrelation experience that evolves at an accelerated speed with the sole purpose to choose more helpful and less aggravating behavior in the interrelation with humans. Several research efforts have aimed to develop flexible architectures for machines that incorporate the understanding of human emotions to make clever decisions and that could make them attractive and accepted by humans. This chapter presents an overview of previous research related to this dissertation as well as the theoretical background of the framework in which it is developed.

2.1 Affective Computing

Important investigations related to multiple intelligences [83], emotional intelligence [84], [85], and emotional neuroscience [86], among others, have revealed that emotion has a strong connection with the decision-making process. Not long ago, the idea of merging computational systems with emotions was formulated [4]; it was named by Rosalind Picard as affective computing [2]. Affective computing, the framework that relates to, arises from, or deliberately influences emotion or other affective phenomena, has two objectives: generation and recognition of human emotions by machines. In relation to this topic, there are important studies that have dealt with this matter, trying to look for the emotional interrelation between humans and machines, specifically the simulation and the interpretation of emotions [3], [53], [87], [88]. A significant number of the studies carried out in affective computing have analyzed the corporal reactions and characteristics from the nature of humans for the purpose of reproducing the characteristics in inanimate entities, or named in another manner, machines. Facial expressions, kinetics of the human body, structural components of the voice or vision, etc., are merely a few examples of outputs with emotional signs that are contained in the natural behavior of human beings. It is very likely, indeed, that a model that replicates intelligent
processes, taking into account the existence of emotional factors, could acquire some level of humanity. Following this road, affective computing heads the list of artificial intelligence-related areas. Many investigations are focused on optimizing the interaction between persons and computers with some aspect of intelligence in the process [31]. As an example, the self-protection and internal regulation that was included in an electronic puppet solved the problem related to saving costs in the film company Stan Winston Studios.

The human body can provide a great variety of information about physiological data; many of these sources belong to hidden affective fields that have not yet been deciphered. This implies that affective computing is constantly growing, taking part in the quest of cognitive measurement approaches of humans’ emotions [2]. While it is true that the sensing and interpretation of emotions by machines would be quite beneficial for groundbreaking research and several application areas, it remains a very difficult task, e.g., factors such as sensitivity and precision are very important. Even if we focus on the complexity of the human nervous system, we can note that signals are unstable from one individual to another, and their analysis will depend on the time of sensing [53]. The objective goal of affective computing is the observation and interpretation of human emotions in order to integrate affective attributes and deal with effective interaction with machines. However, the real problem lies in understand the meaning of the emotional complexity, e.g., mixed feelings can refer to the idea that two basic feelings combine into a compound feeling similar to mixtures of basic colors that produce the sensation of a single new color: “The emotions that result when two or more fundamental emotions are combined, in the same way that red and blue make purple. Judges in these studies have agreed that mixing emotion of love; disgust plus anger produces hatred or hostility” [89].

Affective computing has grown and diversified over the past decades, seeking to resolve the complex riddle of the subjective meaning of “human affect”. At each instant, this technology grows and takes on an ever-increasing role in our lives; maybe it is time that we start building systems that learn to empathize with us, as well. Some initial attempts have explained the level of complexity and the wide range of emotional processes, from the first attempts to decrypt the function of the brain with electroencephalography to an automatic affect sensing, affect synthesis, and the design of emotionally intelligent interfaces with wearable devices. Presently, machines are increasingly approaching more naturalistic human-human interactions, making them an affectively informative environment in which they can display affective states and react to users’ affective states. The great amount of affective data is captured by a diverse group of sensors, such as cameras that can capture facial expressions, ges-
tures, and posture; microphones that record background sounds and speech; neuro-headsets that capture electric signals produced by the brain to detect users thoughts, feelings, and expressions in real time; eye trackers that measure eye positions and movements; and galvanic skin response sensors that measure the electrical conductance of the skin, which can be an indication of psychological or physical arousal, [90], [91], [92].

2.2 The Understanding and Usefulness of Human Emotions within the AI Community

For several years, the Artificial Intelligence community has come to realize that the core to creating affective agents is linked to the understanding of emotional perception in humans. Affective agents with an internal decision system could be capable of choosing an optimum emotional response according to an external cognitive level, by either external or internal emotions captured by them. A considerable number of models, techniques, algorithms, and other tools have tried to achieve a rapprochement between an affective interface and its possibilities to reach a human-centered design; this, however, requires a sound understanding of what emotions in humans involve [93]. So, we consider that an affective agent similar to us is due to emotional skills, and it is possible to think that agents cannot be intelligent without an autoregulation mechanism as a form of emotional awareness, in which the understanding of emotion has an important place in the development. The constant attempts in AI to discover emotional awareness might lead to building agents capable of feeling. Recent studies related to the construction of affective agents have revealed the priority of the design to mimic human behavior with lifelike appearance; even, in some cases, resembling their human creators [9]. However, these entities still lack something in their interaction with humans as they lack humanity, and the emotional factor is very diffuse, e.g., the ability to detect and understand affective states through a human-like face is both an important and necessary step in making systems more agreeable to society.

A significant part of the AI community is mainly interested in humans’ perceptions and explanations of the felt emotions, so that it may be possible to think of methods that can detect information automatically directly from humans and accordingly build affect-based systems. A system that incorporates reason and learning, and is capable of reacting in environments and unexpected situations, will likely have a rule-based system to recognize and solve certain kinds of problems. If affective agents embody real emotional data with a rule-based system, we could use them and modify a character’s behavior [94].
An often overlooked aspect in affect-based systems is related to the real interpretation of emotional expression data in the affective dimension that cannot easily be interpreted by humans. Incorporating affective information into affective agents can potentially enable us to understand the behavior of users, and accordingly facilitate an emotionally smoother and more intuitive interaction with them. The research deployed within this dissertation takes into account a small group of components that constitute an affective agent, as can be seen in Fig. [2.1]. The research is centered on studying the automated analysis of human affective behavior as the perceptual stage of external emotion and could serve to influence the agent’s internal decision as a trigger to an external interaction with a user, e.g., a nursing agent within health systems must not only interact safely with its environment, but it should also act in a way that acquires knowledge about the feelings of patients, and then proceeds according to this knowledge.

![Figure 2.1: Emotional Machine](image)

The field of affective science is interdisciplinary and is growing as never before, as is evident in the number of current scientific efforts. A great part of this research is wholly centered around the capability to detect the emotional state of a person in a natural situation, to subsequently obtain any immediate meaning and use. That is why the experimental stage seeks to deepen our
knowledge of biophysiological signals from humans, e.g., the directional fractionation [95] in which the emotional reaction provides different responses. But even analyzing the patterns of an affective reaction, it is not possible to conclude what kind of emotional reaction has been experienced. Nevertheless, it is possible to observe the produced changes in specific parts of the human body as indicators, e.g., central nervous system, somatic central system, gastrointestinal system [96], cardiovascular reactivity [97], respiratory system [98], etc.

Another important source of emotional information is found in the brain activity. The principal goal is to pursue brain decoding and to reach an understanding about the connection between the physical sensations in our body and our emotions that extends to the neural processes in our brain. All of the experiences carried out in the brain’s emotional processes are focused on the deciphering of basic principles governing brain organization and how it encodes the emotion and the kind of responses that constitute it. Responses include the stereotyped or individual [95] generated models of activation [99], in which the synchrony among the cortical system, the autonomic system, and the somatic system are responsible for the induction of different affective states. The experimental research has used the activation of physiological variables (for instance, increased pulse) and it aims to track a human’s emotional reactions to particular external events [100]. As part of the neurological analysis, the brain cortex and its structures determine the cognitive processes [101], e.g., the reptilian brain is related to certain processes of emotional activation [102]. The brain cortex is an inhibitory variable in affective reactions, and it includes emotional expression, cognitive expression, and decision processes [103]. The frontal lobe and the base of the brain are essential to add emotional value about judgments and decisions [104]. The advancement of these theories evolves in parallel with the advances in biomedicine, which provides tools that make possible the deep analysis of the complex nervous system that takes part in the deployment of an emotion. In neurophysiology, recent techniques provide an electrical record, as real data of cortical areas, the acoustic sensory system, visual, somatosensory, magnetoencephalography, neuroimaging, and computed axial tomography that further reveal more detailed brain electrical activity during signs of emotion, providing leads about the triggers of emotions [105]. For AI, the emotion dimensions can be used to describe general emotional tendencies, including low-intensity emotions. Data could be distinguishable from each other, e.g., changes of dimensions are depicted in facial expressions that reflect directly an emotional state, which entails inducing different affective reactions in the activation of neuronal bases, [106], [107], [108], [109].
Emotions are based on a wide range of seemingly unrelated events, and they do not share any physical features or properties, but all of them can cause the same response [110], [111]. The process of valuation of appraisal is a dynamic process that is precedent to emotions, and it informs us about the situation in which we find the relevance of objectives, goals, and beliefs even if we have sufficient resources for facing it. An appraisal could indicate a momentary affect, averaged to indicate mood over time. These appraisals are combined to produce an overall system equilibrium value using a level evaluation component. The component could be used as an input to an attention stage within an artificially intelligent cognitive system.

Emotions have several levels of variability, e.g., from extreme joy to the total lack of joy in extremely painful situations; this variability is called “intensity”. Emotions can provide interesting information to deal with difficult issues in computational modeling of emotion, functioning as the relation between emotion and mood, the integration of different emotions, emotion dynamics, and the influence of emotion on behavior [112], [113]. Transient emotions with a higher level of intensity use a great part of the frontal region that is activated [114], [115], e.g., a child increases the electrical activity of the brain during maternal separation with notable behavioral signs of discomfort [116]. This can cause the collateral impact of the environment and perceived stimulus, as each action affects the emotional state of individuals at each instant. Emotional intensities could be classified as emotional labels with a great range of possibilities. A level of correlation between all of the emotions are experienced and expressed in a spontaneous manner [117], e.g., the expressions of the face are a clear manifestation of the intensity of an involuntarily emotion without previous organization or intention. The duration and intensity of emotions within negative events are higher than in positive ones [118]. In this case, the impact could be measured by the temporal pattern of emotional responses, which is equivalent to the time that individuals experience as intensity (in positive emotions, the time is approximately 40 minutes, and for negative ones, 110 minutes). If there is a possibility to quantify the emotions, it is also suitable to establish “laws of emotions”, because it is possible to follow its behavior [119].

The idea to generate rational actions using human emotion as a reference in computational models in AI has become quite fertile, because there is a shared language of utility, probability, and reasoning about others, taking into account the measures of expectations, alternatives of situational states, and motive-consistency of events. Some approaches of certain representative models can be found in Decision Affect Theory (DAT) [120], FLAME [121], and a Computational Model of the Appraisal Process (EMA) [122]. The models
work with representations of future states to enhance anticipatory competencies. They refer to some of the several functions that expectations can embed in a system guided by goals, e.g., a clear representation is the Roboceptionist model. This functions as a simple generic model based on discrete emotions, attitudes, and influences that exist among them. The Roboceptionist’s mood does not change just in the current state; at the same time, it is capable of changing memories that it had about past events and the expectations of future events \[123\]. Expected utility \[124\] in decision-theoretic planning provides a useful framework for the decision-making process influenced by emotions, determining the utility of a possible future state and thus the expected utility of an option. But talking about emotions in this framework involves the representation but not the acquisition of an emotional process, e.g., a strong negative variable of emotion associated with one option can affect the expected utility of another option, so that it is not an optimal analysis; to put it another way, it is a representation of emotions to influence a decision. Individuals usually employ several inputs for decision-making. This risks confusion due to the large set of variables, with emotions being one of them, possibly neglecting the emotions and behaviors that will alter a decision dramatically, causing different consequences. The effects of emotions in the subjective expected utility theory is shown through the connection with decision-making under risk. Some part of this decisional process has a significant emotional charge, specifically, when particular activities are undertaken with a large amount of emotional correlates within risk-averse decisions.

AI systems interacting with people as affective agents need to have appropriate emotions or, minimally, an understanding, even if they do not actually have them. Such research is ultimately focused on proposing detailed models for explaining the cognitive, perceptual, and emotional qualities of humans and the advantages that would be associated with possessing that knowledge. There are also an increasing number of examples based on the simulation and acquisition of human emotional control. The next group of examples is by no means intended to give a complete overview of the present affective systems concerned with emotions in AI, but all of them have tried to construct a path to the usefulness of representing human emotions. Fuzzy Logic Adaptive Model of Emotions (FLAME) \[121\] is based on fuzzy logic to represent emotions, events, and emotional observations. It includes several inductive learning algorithms for learning patterns of events, associations among objects, and expectations. It does not entail a direct use of human emotion, but merely an approximate representation, in which the quantification of emotion intensity is calculated, taking into account the value of an expectation and the measurement of the convenience of an event. The in-
tensities can be expressed through: 

\[ \text{Joy} = (1.7 \times \text{expectation}^{0.5}) + (-0.7 \times \text{desirability}), \]

\[ \text{Sadness} = (2 \times \text{expectation}^2) - \text{desirability}, \]

\[ \text{Disappointment} = \text{Hope} \times \text{desirability}, \]

\[ \text{Relief} = \text{Fear} \times \text{desirability}, \]

\[ \text{Fear} = (2 \times \text{expectation}^2) - \text{desirability} \]

\[ \text{and } \text{Hope} = (1.7 \times \text{expectation}^{0.5}) + (-0.7 \times \text{desirability}). \]

The model of Cañamero represents a two-dimensional environment [125]. An agent’s behavior is driven by motivational states and a set of basic discrete emotions (anger, boredom, fear, happiness, sadness, interest) as alert mechanisms. The model handles two variables related to valence and physiological activity. A behavior can be activated and executed with different intensities, which depend on the motivations related to them [126]. Motivations are modeled through the equation, \( m_i = d_i + (d_i \times \alpha c_i), \) where \( d_i \) is the psychological deficit and \( c_i \) are the external stimuli that allow executing behaviors to achieve objectives and satisfy needs. The parameter \( \alpha \) will determine the quality of a stimulus, e.g., hormonal levels and corporeal states. The cathexis model [127] generates emotions and influences in an agent’s behavior. Emotions, moods, and temperaments are modeled as a network composed of special emotional systems. The network is comparable to Minsky’s “proto-specialist” agents network [30]. The cathexis approach takes emotions as drives of different emotional triggers. The triggers can be natural signals as a result of outputs from a set of sensors or chemical signals issued by the brain. Emotional responses also depend on probabilities and unobtained outcomes. Unexpected outcomes have greater emotional impact than expected outcomes [120]. Decision Affect Theory (DAT) translates risk, probability, and expectations of events into a one-dimensional feeling. It might be compared to the valence dimension of Russell [128]. This model is supported on theories of disappointment and regret, and it tries to capture the effect of other alternatives on the experienced benefit of an outcome. The EMA model is based on the theory of appraisal of emotions and simulates various behaviors in autonomous agents [122]. This model works oppositely to appraisal models and highlights the aspect of coping. The model includes a level of appraisal-derivation that interprets a representation of the person-environment toward a set of appraisal variables, a level of emotion-derivation that takes the appraisals and produces an emotional response and a set of coping strategies, triggered because of emotion, that subsequently manipulate the person-environment. In the case of appraisal-based architectures directly extracted from the behavior of emotions, the Affective Reasoner [129] considers the social interaction, taking into account the process of modeling human emotions. It is widely used to engender emotional competence in affective agents including some reified representations of a finite number of discrete emotional states through which all emotional processing is explicitly routed. The core is based on the OCC Model of Emotions in Embodied Characters [130], which assesses the relationship between events and
an agent personality, described by its goals, social standards, and preferences. Taking into account one part of the ideas and considerations deployed in this dissertation, all of them are centered around using human emotion as a data indicator to pursue a future action. We can locate our research within the group of all efforts that pursue the understanding of emotions and use this information provided to act according to the appropriate emotional behavior of a human. This is indeed an area of growing interest supported in the technological revolution that will allow the interpretation of human emotion like never before [131]. It is not about the exhaustive understanding of the emotion in itself; rather, it is a step closer to attaining a decision in an external affective system with restrictive emotional connotation limited by humans. Affective systems are taking on human tendencies as emotional understanding and, as a result, our relationship with technology is expanding. As we continue to strive for validation from our devices and systems, the emotional interplay between man and his man-made objects can only be achieved with the very real power of AI technology to focus a significant part of these architectures that cater to emotional aspects.

2.3 Required Backing to Pursuit of a “Decision with Emotional Connotation”

2.3.1 Adversarial Risk Analysis

Adversarial Risk Analysis (ARA) is a relatively new approach that was established because of the several attacks from the Islamic extremist group al-Qaeda against targets in the United States, also called the 9/11 attacks, in which there was massive investment, by many countries, in measures to reduce vulnerability. Several alternatives were proposed to deal with the cost-effective prevention of a terrorist attack, but without any methodology for deciding which security measures should be adopted, e.g., alternatives as large investments to provide conceivable defense strategies based on a game-theoretic perspective, and probabilistic risk analysis, among others. Given these challenges, ARA was proposed to deal with threats originating from intentional actions from imminent adversaries. ARA combines statistical risk analysis and game theory to create new methods for the analysis of decision-making, and it has motivated extensive research within scenarios such as counter-terrorism, cybersecurity, and competitive corporate decision-making [132], [133], [134], [135].

Adversarial Risk Analysis is concerned with problems in which there are two or more intelligent opponents who make decisions with uncertain outcomes...
ARA is motivated by applications in competitive decision-making, which is housed within the perspective of non-cooperative game theory, e.g., \[136\], \[137\], and it has seen a renewed interest in developing practical tools and theory to analyze the strategic calculations of intelligent opponents, the vast majority of them acting in scenarios with random outcomes. ARA has a clear game theoretic flavor, although supported with some decision analytic based approaches \[138\], \[139\].

The core of ARA uses a Bayesian model for the decision-making processes of one’s opponents to develop a subjective distribution over their actions. The strategy enables the application of traditional risk analysis to maximize the expected utility. A player acts in order to maximize its expected utility under subjective beliefs about the probabilities of its opponents’ actions and utility functions. The ARA formulation is fully Bayesian, and its implementation enables great flexibility in tailoring the analysis to the situation and the counting of different kinds of uncertainty that arise. The problem analyzed with ARA involves a hierarchy of nested models for the decision processes that imply several steps of decision analyses; as in chess, the number of levels typically correspond to the number of moves ahead that someone thinks. This subjective belief is developed using a hierarchy following a Bayesian version of level-\(k\) thinking \[140\], \[141\]. The \(k\)-level hierarchy indicates how deep the player thinks its opponents’ decision-making process are. The absence of any strategy followed by the opponents would determine a zero-level model. The first-order analysis model will be addressed if the opponents look for modeling of the player’s thinking. Thus, introducing a second-order analysis model is posed if the opponents model the player’s model of the opponents’ decision-making, and so on. Other approaches that require the same Bayesian strategy result are likely to have obstacles to find the correct mechanism that allows a decision-maker to develop subjective probability distributions that precisely delineate an opponent’s behavior \[142\], \[138\]. ARA solves the problem using a “mirroring” procedure, in which the decision-maker imitates the opponent’s analysis, taking into account the fact that the opponent may perform simultaneously an analysis of the decision-maker’s process. One obtains a probability distribution over the opponent’s options. This strategy allows traditional risk analysis to derive the action that maximizes expected utility \[132\].

Within a scenario are found a decision-maker \(A\) and an adversary \(D\); these two entities interact among themselves. In general, suppose that \(D\) can select actions from the set \(D = \{d_1,\ldots,d_m\}\) and \(A\) can select actions from the set \(A = \{a_1,\ldots,a_n\}\). Both of them also have utility functions \((u_A(.), u_D(.))\), a collection of probabilities about outcomes \((P_A, P_D)\) and expected utilities \((\psi_A,\))
$\psi_D$). The utility that $A$ expects from performing action $a \in A$ when $D$ choose action $d \in D$ is represented through

$$\psi_A(a, d) = \int u_A(c) \pi_A(c \mid a, d) dc,$$

where $\psi_A(c \mid a, d) \in P_A$ models $A$’s beliefs about the consequences ($c \in C$) for the pair of actions $(a, d)$. The utility that $D$ expects from performing action $\psi_D(a, d)$ is assessed as $A$. In a game-theoretic non-cooperative scenario, the aim is to establish a Nash Equilibrium joint plan out of the individual plans of $A$ and $D$. Here, to find the Nash Equilibrium, $(a^* \text{ and } d^*)$ is used

$$\psi_A(a^*, d^*) \geq \psi_A(a, d^*) \forall a \in A,$$

$$\psi_D(a^*, d^*) \geq \psi_A(a^*, d) \forall d \in D.$$

Let us suppose that one of the players $D$ receives the full support against the player $A$ in a simultaneous decision game. In that case, the solution will be as follows:

$$d^* = \arg\max_{d \in D} \sum_{a \in A} \left[ \sum_{s \in S} u_D(d, s) p_D(S = s \mid a, d) \right] \pi_D(A = a).$$

In the solution, $S$ is an uncertain outcome that approximately represents the success of $A$, and $p_D(S = s \mid a, d)$ describes the $D$’s beliefs about $A$’s success given $A$’s and $D$’s actions. The main difficulty is the assessment of $\pi_D(A = a)$, for accomplishing that purpose $D$ would solve $A$’s decision analysis,

$$a^* = \arg\max_{a \in A} \sum_{d \in D} \left[ \sum_{s \in S} u_A(a, s) p_A(S = s \mid a, d) \right] \pi_A(D = d).$$

Using a Bayesian strategy, we should put a distribution over $A$’s $(u_A, p_A, \pi_A) \sim (U_A, P_A, \Pi_A)$, and this may be estimated by Monte Carlo simulation [132]. Taking into account that $A$ is a rational player, $D$ would model $A$’s decision problem; thus, $D$ shall assume that

$$A \mid D \sim \arg\max_{a \in A} \sum_{d \in D} \left[ \sum_{s \in S} U_A(a, s) P_A(S = s \mid a, d) \right] \Pi_A(D = d).$$

The elicitation of $\Pi_A(D = d)$ should require a deep analysis for the next level of hierarchical thinking. Assuming that $D$ considers that $A$ is assessing $D$ as a strategic thinker, she would solve,
\[ D \mid A^1 \sim \arg\max_{d \in D} \sum_{a \in A} \left[ \sum_{s \in S} U_d(d, s) P_d(S = s \mid a, d) \right] \Pi_D(A^1 = a). \]

This would lead to an infinite cascade of recurring thoughts about what the other is thinking about. According to this, the process should stop when we no longer have more information about utilities and probabilities at some level of the recursive analysis.

2.3.2 Emotion Recognition by Machine Learning

**Emotion Recognition in Facial Expressions**

Identifying emotions through the gestural information in a face is a normal task for humans. The human face has many indicators that conduct dynamic information about various subtle emotional cues. Machines are still sometimes seen as a union of cold components that preclude the understanding of a human’s emotional state. The development of machines capable of reproducing the vision sense in face perception has benefited areas as varied as computer science, cognitive science, mathematics, physics, psychology, and neurobiology. We may even think that a machine with sophisticated and expensive vision equipment is needed for decrypting emotions, in some cases surpassing human levels of recognition, but compared with the human sense, it looks to be more complex. The human face acts clearly as the display board on which emotions and intentions show their variability.

In a groundbreaking investigation, the underlying concept of facial expression recognition was developed in [143], with an early attempt to automatically analyze facial expressions by tracking the motion spots on an image sequence. There exists a correlation between all of the emotions that are experienced and expressed in a spontaneous manner [117], e.g., the expressions of the face are a clear manifestation of the intensity of an involuntary emotion, without previous planning or intention. Exploring new ways in which machines can understand the defined meaning of gestures and putting them into the context of the feelings with which they are expressed, should be the key in a closed loop with humans. Following the same line, several approaches to decrypt emotions from the face have been reported. These approaches have attempted to tackle the group of basic emotions based on simple static models, with successful results [144]. From this point of view, this group of studies also analyses facial emotional expression from diverse physical zones, in tune with mechanical movements of the facial muscles. During the decryption quest of more subtle
emotional expressions that are carried out over the face, it has been found that the dynamic information has been important for detecting emotions. The analysis is based on more natural sequences of facial expressions rather than the isolated captures usually depicted in early databases [145], [146]. The emotion detection applied over natural sequences is more difficult to achieve than when it is applied to isolated expressions [147].

Another road to categorizing information from facial expressions is based on an explicit coding system related to their dynamics. The facial movements are coded in a set of action units (AUs) with a muscular basis named the Facial Action Coding System (FACS) [148]. Several authors have worked on this research; taking into account the dynamics of changing faces, the automatic captures were covered from action units from facial motion sequences [149], [150]. The latest developments have applied facial tracking and recognition techniques supported by advanced computational techniques. The issue of face tracking has also been the topic of several face-based emotion recognition systems. Various methods include the measurement of optical flow [151], active contour models [152], face identification, recovery of face pose and facial expression [153], a probabilistic approach to detecting and tracking human faces [154], active [155] and adaptive [156] appearance models, multiple eigenspace-based methods [157], tracking by a robust appearance filter [158], and facial motion extraction based on optical flow [159].

The machine learning framework shows several classifiers used in various tasks related to facial expression recognition. Each classifier has advantages and disadvantages in order to deal with the recognition problem. These classifiers include Support Vector Machines [160], Bayesian Network Classifiers [161], Linear Discriminant Analysis [162], Hidden Markov Models [163], and Neural Networks [164]; additional detail is provided in [165]. We could highlight some approaches within the facial expression of emotion: parametric models to extract the shape and movements of the mouth, eye, and eyebrows have been used in the works of [166]; major directions of specific facial muscles were the input in an emotion recognition system treated in [167]; and permanent and transient facial features such as lip, nasolabial furrow, and wrinkles were taken as recurrent indicators over the emotions [168]. We can also point out geometrical models that heavily use the locating rectangles containing the appropriate muscles of the face and are very useful for synthesizing facial expressions. In order to extract the shape and movements of the mouth, eye, and eyebrows, parametric models were used in the works of [169]. The major directions of specific facial muscles were the input in an emotion recognition system treated in [170]. The permanent and transient facial features
such as lip, nasolabial furrow, and wrinkles are useful recurrent indicators of emotions [171], but in this technique, it is important to use geometrical models that locate the shapes and appearances of these features with great accuracy; the presence of the features and their geometrical relationships with each other appear to be more important than the details of the features [172], [173].

Nonetheless, the common denominator in facial emotion detection is that it is always initiated with a detection of face zone, and involves extraction and tracking of relevant facial information. Finally, the facial expression classification occurs. Then, with all of this information, the system proceeds to analyze the facial expressions to estimate emotion-related activities.

Facial Emotion Recognition through Neural Network-Based Methods

The extraction of emotion from the static image allows the recognition of several physical features such as eyes, wrinkles on the forehead, size of eyebrows, color of the skin, etc., and their corresponding sizing and location. In this case, the neural network is accurate for the acquisition of nonlinear mapping between different sets of data. This analysis allows for decoding the relationship between the physical features of a face and its impression. The potential of the neural network-based methods is the performance of facial expression classification into a single basic emotion category. The sort of selection of six basic emotions using neural network-based methods was proposed in the work of [173], where the units of the input to the ANN correspond to the brightness distribution data extracted from an input static image. The average recognition rate was 85% in a group of 90 tested images. Furthermore, the ANN can be reinforced with the use of a hybrid approach. In the work of [174], the ANNs are often combined with Hidden Markov Models (HMMs) and are employed in facial emotion classification. This analysis used an ANN to estimate the posterior for the discriminant HMM, and it achieved positive results on the recognition of emotion in the upper and lower parts of the static image separately. The research of [36] used the analysis of principal facial regions employing principal component analysis and neural networks. The classifiers were built in a group of 15 neural networks. Only one ANN in this group was used for region detection, and the other 14 were used to learn to recognize seven universal emotions over the eye and mouth regions. The conducted experiments showed a 46% to 80% rate of successful recognition that had an average precision of 70%.

Positive results in the emotional classification of an input static facial image can also be found in [175]. This study showed outputs from six differ-
ent classes of neutral emotions. In the construction of the ANN, the output layer had contained seven units, each of which corresponds to one category of emotion; the average of the correct recognition rate achieved was 86%. In the works of [176], the neural network performs a nonlinear reduction of the dimensionality in the input image, because the data of interest lie on an embedded non-linear manifold of the higher-dimensional space. In this step, the algorithm makes a statistical decision about the category of the observed expression. The set of outputs gives an estimation of the probability of the examined expression belonging to the associated category. The power of this classification achieved a 90.1% recognition rate.

Other types of approaches as in the work of [177] propose the use of Multilayer Feed-forward Neural Networks and a Radial Basis Function Networks, commonly used in nonlinear mapping approximation and pattern recognition. The experiment shows the classification of seven basic types of emotions: Neutral, Happiness, Sadness, Anger, Fear, Surprise, and Disgust. The Euclidean distances from the contour point in the static image and the geometric coordinates from facial characteristics points represent the set of data input into the neural network. This approach was tested with the set of images from the Japanese Female Facial Expression Database (JAFFE), and it reached 73% accuracy. Projections of feature regions on a fixed filter set of images were proposed in [178]. This model used a feed-forward neural network in which the inputs are the group of features based on representation of the face, considering the observations found in the study of human expressions. The network model consists of ensembles of 11 feed-forward, fully connected common neural networks. This architecture has 105 inputs per network, and each network includes hidden layers with 10 nodes. The training of each network works with online backpropagation. The outputs of each network are combined to produce a percentage value for the classification of each emotion. Experiments conducted in [179] showed the applicability of neuro-fuzzy networks to extract emotion in facial motion. This approach has attempted to classify primary and intermediate emotions using the definition of Facial Animation Parameter (FAP) intensities and the definition of emotions in terms of FAPs. The FAPs specified in MPEG-4 compose a very rich set of parameters that allow a wide range of facial motions. The challenge of this classification is performed by translating the Feature Point (FP) movements into the FAPs.

**Emotion Recognition using Body Pose**

Bodily expressions convey important affective information, although this modality has been relatively neglected in the literature as compared with facial ex-
pressions and speech, and has been a major challenge over the past several years. During this quest, several misclassifications related to interpretations have been found, often due to the structure of the human body. In order to obtain an accurate emotion, the use of a multimodal framework between speech and the body could replace the wrong data [180], as the degrees of freedom of the human body are higher than those in the face alone, and its overall shape varies strongly during articulated motion.

In machine learning research, recent results related to object recognition have shown that even for highly variable visual stimuli, quite reliable categorical decisions can be made from dense low-level visual cues [181]. Many researchers have collected lists of stereotypical features (see [182]) in order to decrypt emotional bodily expressions, whereas others have argued for diverse patterns along a number of more abstract dimensions, using terms like force, speed, energy, directness, etc., [183], [184]. Despite the usefulness of these features, based on generalities, they tend to focus on dynamic properties of bodies, and usually, they fail to make clear predictions regarding the bodily poses that may be associated with different emotional states, e.g., likely configurations of the distribution of head, trunk, arms, shoulders, and legs in an image could be of noticeably grainy appearance and difficult to predict. Nonetheless, there exists a variety of sources that offer more or less detailed descriptions of emotional bodily poses [185], [186], [187].

Facial gestures, kinetics of the human body, and bodily poses can be underlined as clear indicators of emotional states [188]; some of these indicators are critical in emotional recognition from affective states [44]. The body movement of humans differs from other emotional indicators like speech and face gestures, since it is the only visual stimulus that we can perceive and produce with several degrees of freedom, with several combinations from all components of the human body [189]. Expressive body movements are strongly influenced by emotions and movement qualities highlighted in [190], [182] and [183]. Several experiments with actors who express emotions through body posture were analyzed in [42] by using photos without tridimensional information; the set of photos was decoded using low-level visual data.

Some experiments have focused on body posture representation; a system can capture different body positions and generate a set of features, e.g. distance and angles between shoulder and head [43]. In addition to furthering basic understanding of human behavior, work on biological movement can reveal how stimuli act as triggers, leading to better design of computational models of emotion, focusing only on visual analysis of affective body language [191].
The human brain can generate empathic connections at a social level; for this to be possible, the motion-visual neurons are affected biologically by bodily motion in many visual areas. This may allow a deeper understanding about the comprehension of others humans and how emotions are key to studying empathy [192].

**Machine Learning in Bodily Emotion Detection**

The first research in automated emotional bodily recognition was reported in [193]. This work focused on posture analysis through Tekscan’s Body Pressure Measurement System (BPMS). The system works in a learning environment, sensing the temporal transitions of posture of children. A neural network provided real-time classification of nine static postures with an overall accuracy of 87.6%. Machine learning algorithms were used in order to detect boredom, engagement/flow, confusion, frustration, and delight, by kinetics of movements of students during a learning task [194]. Two sets of features were selected from the pressure maps that were automatically computed with the BPMS. The emotion detection showed accuracy in emotions like boredom, confusion, delight, flow, and frustration, differentiated from neutral, with 73, 72, 70, 83, and 74 % accuracy, respectively.

In [41] and [195], a real-time analysis of expressive gesture in full-body human movement was performed based on computer vision algorithms, where the quantity of motion and contraction index of the upper body as well as velocity, acceleration, and fluidity of limbs and head were measured. The Bayesian network-based classifier achieves a correct recognition rate of 61 % from four emotions: anger, joy, pleasure, and sadness. The affective states using information of facial expressions and upper-body gestures were decrypted in [196]. Expressions like anger, anxiety, disgust, happiness, and uncertainty are recognized with a performance of 90 % accuracy by using body expressions only and by employing a Bayes network. The kinetics of four emotional states was recorded using a 3D motion capture system, and with these data, several classifiers were tested and trained. Neural Networks and SVM were able to achieve a correct recognition rate of 84 % [197].

Similar experiments in [198] were developed using the data from records of a standard digital video (DV) camera. The expressive motions were distinguishable by a Bayesian network-based classifier at a rate of 90 % accuracy. The streams of 3D measurements were the input data to binary SVM classifiers [199]; the experiments covered the multi-class classification of six emotions based on a combination of classifiers that used gesture segmentation by ki-
netic energy. The reason to use SVM classifiers was that the descriptors were easily separable. A pattern recognition problem was addressed in [200]; the classification used suitable features from gestures provided by a Kinect sensor. The joints from the upper body were coded using the angles and positions. The feature spaces obtained were classified using a number of different classifiers. The average classification obtained for binary decision tree, ensemble tree, k-NN, SVM with radial basis function kernel, and neural network classifier with backpropagation learning were 76.63 %, 90.83 %, 86.77 %, 87.74 %, and 89.26 %, respectively. Emotions like disgust, fear, happiness, surprise, sadness and anger, anxiety, boredom, puzzlement, and uncertainty were decrypted from The Bimodal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior (FABO) [201]. The problem of multi-class classification was solved using an SVM with a RBF kernel as multi-class classifier. The average accuracy achieved by the threefold cross-validation was 83.1 %. In [202], a Random Forests classifier faced the problem of multivariable time series classification of extracted features from psychological experiments [182]. Features like low-level postural, high-level kinematic, and geometric were calculated as well as statistical cues. The experiments over the real-time expressive gesture recognition system achieved an overall recognition rate of 75.41 % (138 correctly classified out of 183).

**Emotion Recognition Using Speech Cues**

The emotional decryption of human speech has been a challenging research issue for many years and one with growing importance, with many publications on the topic [203]. A growing scientific field in which it is crucial to have a robust speech emotion recognition system is that which comprises medical robotics, because of the emotional knowledge that the robots could learn and acquire in social environments and employ to take care of human beings [93].

The applications for emotion detection in speech are varied with practical uses, taking into account that all of the efforts are focused on the decryption of the emotional meaning from all structures of speech prosody. Research in acoustics has found connections between emotional cues provided through the spoken dialogue being of potential value to learn about human emotions [204]. Theoretically after the decryption of emotional cues, it is possible to recognize the emotions present in speech utterances, bearing in mind that the emotional content of speech is not influenced by the speaker or the lexical content.

The emotional decryption of speech has a multitude of features that have been consistently used in much research [205], but a rule to follow in order to man-
of the multiple features has not yet been established. Studies based on psychological and neurobiological factors have discovered how prosodic cues are characterized by fundamental frequencies and how the intensity of the voice can show variable levels across different speakers [206], [207]. Emotional indicators are located along the speech chain in short-term spectral features and sound quality cues [208], [209].

Prosodic features like pitch, loudness, speaking rate, duration, pause, and rhythm show strong correlations between them, providing valuable emotional information; they have been viewed as the most informative group that supplies many emotional recognition cues [210]. Taking into account the analysis of the entire segment of voice, statistical functions like mean, median, minimum, maximum, standard deviation, or more seldom, third or fourth standardized moment, are applied to the fundamental frequency (F0) base contour [13], [211], [212]. Human speech contains spectral features including Mel Frequency Cepstral Coefficients (MFCCs), generally used in speech recognition with great accuracy in emotion detection [213], [214]. Features such as Predictive Cepstral Coefficients (LPCC) or Mel Filter Bank (MFB) have been commonly used to classify and depict emotions [215]. Parameterization techniques such as MFCCs and Relative Spectral Transform - Perceptual Linear Prediction (RASTA-PLP) work with the same performance and are commonly used to extract features from speech; many of their parameters are considered within the nature of speech [216].

The voice quality analysis from a phonetic perspective [217], [218] and linguistic features show strong correlation to exist between voice, pronounced words, and emotions [219], [220], [221]. For example, different levels of voice could be depicted by neutral, whispery, breathy, creaky, and harsh or falsetto voice. The emotional features located in chains of words are depicted by the affective states associated with each specific word; many of them are related to the probability of one emotion within a certain sequence of words [222].

**Emotion Recognition Using Audio-Visual Cues**

Humans use different sources to acquire knowledge about certain emotions from other humans whom they constantly face. The entire emotional decryption process will be part of a non-verbal communication; the mix of sounds and images can serve as emotional cues. The information fusion that humans collect between facial gestures and distinguishing voice signs may contain indicators about the emotion depicted at the precise moment, consciously or unconsciously, carrying specific messages [223]. The visual markers of the face...
could be influenced by emotional indicators in speech; this aspect would ensure a global impression formed about the emotion communicated between sensory modalities. The human social cues serve several purposes in social interactions. Among these cues, speech prosody is known to play an important role in social communication, facilitating the understanding of speakers’ intentions during interpersonal communication [224]. Emotional cues such as audio and face images can be concatenated to construct joint feature vectors that are then used to develop an emotional automatic classifier [225]. Some approaches to automatically recognize emotions are improved through unimodal systems. The outputs from the unimodal systems could be fused at the decision-level, increasing the accuracy of emotional detection [226], [227]. The features used for speech were prosodic features such as characteristics of the fundamental frequency (F0). In the case of video frames, differences of movements and geometrical distances between six different zones in a face were employed; the data encrypted in audio-visual features had been used to resolve inaccuracies between the confusing emotion classes. A rule-based approach was taken into account to classify the classes of emotion; in these classes, one feature at a time was selected. This technique of classification achieved the ability to detect and discriminate finer details. A thermal distribution over a human face was used as an additional source in a multimodal system. Data for speech, voice, and thermal levels in skin were discriminated at the decision-level using empirically determined weights. The union of three modalities showed more accuracy that using only two [228]. Differences in audio-visual emotion recognition between spontaneous displays (natural or unintentional) [229] and deliberate displays (posed or intentional) [230] are important to understand of the nature and role of spontaneous human expression in social interaction. Spontaneous emotions send signals not included in the criteria of social norms that pursue the goals for communication. To decrypt the emotions in audio-visual data, three levels of fusion are typically used: feature-level [229], decision-level [231], and model-level fusion [232]. The first one constructs joint feature vectors with the concatenation of prosodic features and facial features to build an emotional classifier. The second one combines in a total result the outputs originating from each modality. Finally, the third one is a hybrid of the first two; it aims to combine the benefits of both the feature-level and decision-level fusion methods.

A wide variety of schemes for audiovisual emotion recognition have been reported as standard methods for related tasks supported by machine learning techniques [13]; many of them employ the auditory and visual features discriminated by classifiers because of their excellent performance. Some of the innovative investigations that had been joined to the discrimination
of emotional cues based on machine learning classifiers are: Bayesian classifiers [180], Decision Trees [233], Support Vector Machines (SVM) [234], [235], [236], Artificial Neural Networks [237], [238], [239], K-Nearest Neighbor [240], Bayesian Networks [241], Hidden Markov Models (HMM) [163], [242], AdaBoost [243], and Genetic Algorithms [244].

2.4 Classifier Methods and Performance Measures

2.4.1 Classifier Methods

Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a supervised learning method that can be applied to classification or regression; this method originated in statistical learning theory. SVM is an extension to nonlinear models of the generalized portrait algorithm developed in [245], offering robust classification to a very large number of variables and small samples. SVM is capable of learning complex data from classification models applying mathematical principles to avoid overfitting. There are a number of kernels that can be used in Support Vector Machines models. These include linear, polynomial, radial basis function (RBF), and sigmoid, with the first two being the most used.

k-Nearest Neighbors (kNN)

k-Nearest Neighbors (kNN) is one of the simplest classification algorithms available for supervised learning used in statistical estimation and pattern recognition that has been employed from the early 1970s as a non-parametric technique. The kNN stores all available cases and classifies new cases based on a similarity measure (e.g., Euclidean, weighted cosine distance, etc.). Also, it can be explained as a lazy learning method that searches the closest unlabeled examples of the test data in the feature space, based on distance function [246].

Multilayer Perceptron (MLP)

Within the connectionist techniques is also found the Artificial Neural Network (ANN). A Multilayer Perceptron (MLP) is a feed-forward ANN more used for classifications. The most used training algorithm in MLP is Backpropagation [247]. The learning process follows two steps: the first is a forward processing of input data through the neurons that produces a forecasted output, and the second is the adjustment of weights within the neuron layers to minimize the errors of the forecasted solution compared with the correct output.
BayesNet (Bayesian Network)

BayesNet (Bayesian Network) is a graphical model (GM) for probabilistic relationships among a set of variables; it is used to represent the uncertainty of knowledge [248]. The graph depicted in bayes Net is composed of nodes that represent propositional variables of interest and links that represent probabilistic dependencies among the corresponding random variables. For each node there is a probability table specifying the conditional distribution of the variable given the values of its parents in the network. The network supports the computation of the probabilities of any subset of variables given evidence about any other subset. These conditional dependencies in the graph are generally calculated by using known statistical and computational methods.

Decision Tree

The Decision Tree is a powerful tool for classification and prediction that makes possible the representation of rules. The group of classification rules is simple and easy to understand. The rules represent the information in a tree based in a set of features. A classic decision tree, named Iterative Dichotomiser 3 (ID3), is a basic technique to construct a decision tree based on information gain/entropy; it is established on growing and pruning [249]. Another topdown decision tree inducer for continuous values is C45 [250]. C45 is named as J48 in WEKA [251], and it uses the information gain as a measure to select and split the nodes.

Naive Bayes classifier

The naive Bayes classifier is a supervised learning method as well as a statistical method for classification [252]. The probabilistic classifier is based on the well-known Bayes theorem with strong assumptions; it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outputs. One of the advantages is the robustness to noise in input data. The classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable.

2.4.2 Performance Measures

Evaluation measures play an important role in machine learning in order to evaluate the performance of classifiers, which are principally focused on handling two-class learning problems. This research has addressed a classification problem of six classes formed by six universal emotions.
Confusion Matrix

The confusion matrix is a measure that contains predicted and actual information about the classification performed by a classification system. The vast quantity of measures used in the performance evaluation of binary problems could also apply to multi-class problems. The performance level of classifiers can be assessed with a $m \times m$ confusion matrix in a problem with $m$ classes, as depicted in Table 2.1. The rows that describe the matrix show the actual classes, while the columns are the predicted classes.

<table>
<thead>
<tr>
<th>Predicted Class$_1$</th>
<th>$\cdots$</th>
<th>Predicted Class$_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Class$_1$</td>
<td>$CM_{11}$</td>
<td>$\cdots$</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>$CM_{1m}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\ddots$</td>
</tr>
<tr>
<td>True Class$_m$</td>
<td>$CM_{m1}$</td>
<td>$\cdots$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$CM_{mm}$</td>
</tr>
</tbody>
</table>

Table 2.1: Confusion Matrix

Accuracy

Accuracy is defined as the percentage of correctly classified cases of the dataset. Based on the confusion matrix, the accuracy can be computed as the sum of the main values from the diagonal of the matrix, which represents the correctly classified cases divided by the total number of instances in the dataset, as shown in Eq. 2.1.

$$\text{Accuracy} = \frac{\sum_{i=1}^{m} CM_{ii}}{\sum_{i=1}^{m} \sum_{j=1}^{m} CM_{ij}} \quad (2.1)$$

where $CM_{ij}$ represents the elements in the row $i$ and column $j$ of the confusion matrix.

Recall

Some measures like accuracy do not represent the reality of the number of cases correctly classified for each class. In order to achieve a deeper analysis, the measure of recall has been calculated for each class. Recall of a class $i$ refers to the percentage of correctly classified cases based on the overall number of cases within class $i$. Eq. 2.2 represents the recall for class $i$.

$$\text{Recall}_i = \frac{CM_{ii}}{\sum_{j=1}^{m} CM_{ij}} \quad (2.2)$$
Precision

Precision or positive predicted value denotes the proportion of predicted positive cases that are correctly classified. Precision is a measure of accuracy of predicted positives cases, in contrast with the rate of discovery of real positives. This measure assesses the predictive power of the algorithm. The precision for class is calculated using Eq. 2.3.

\[
\text{Precision}_i = \frac{C M_{ii}}{\sum_{j=1}^{m} C_{ji}}
\]  
(2.3)

F-measure

The F-measure, also referred to as the F-score or F1 score, is defined as the harmonic mean of precision and recall. It is null whenever one of the two indices is null; the F-Measure increases proportionally when the values of precision and recall increase. A high value of F-Measure indicates that the model performs better on the positive class. The F-measure for class formula is represented in Eq. 2.4.

\[
F_i = 2 \frac{\text{precision}_i \times \text{recall}_i}{\text{precision}_i + \text{recall}_i}
\]  
(2.4)

k-fold cross-validation

A k-fold cross-validation with \(k = 10\) was used to make validations over the classifiers. This technique allowed the evaluation of the model facing an unknown dataset. The group of data is randomly divided into \(k\) equal parts; one part of the group is used as a validation set and the rest \(k - 1\) will serve as the training set. The process is repeated \(k\) times using a different group as a validation set each time; this process continues until each group has been used once as a validation test. Then, the \(k\) results obtained by groups can be averaged to a single result. The advantage of 10-fold cross-validation is that all examples of the database are used for both training and testing stages [254].

2.4.3 Multi-classifiers

A multi-classifier is an ensemble of individual classifiers to build a consistently more accurate classifier than a single classifier. The main goal of these methods is to increase the efficiency and accuracy of classification. The output of the individual classifiers are a mix to yield a final decision over its classification. The combination of outputs is divided in two: selection and fusion. The selection technique uses the output of the “best classifier” as the general output.
of the ensemble, whereas the fusion technique employs a strategy to mix the different results. The literature describes a significant number of methods of fusion in multi-classifier models [255]. The multi-classifier system requires a base of classifiers with different decision boundaries, in order to handle diverse classifiers with different misclassified regions. The diversity of outputs of a multi-classifier is a vital requisite to reach its objective. It is possible to achieve diversity classifiers through different methods. One approach is to use different training datasets to train individual classifiers, requiring the performance of each classifier. Bootstrap aggregating [256] and boosting [257] are the most well-known techniques in this category. Diversity can also be obtained by using different base classifiers such as Vote [258], Stacking [259] and Mixture of Experts [260]. On the other hand, the use of different features in the data training is another method to achieve diverse base classifiers, including the random subspace method [261] and more recently [262].

2.5 Summary

This dissertation derives its inspiration from various disciplines and frameworks, including affective computing, physiology, machine learning, affective decision making, and computer science. This chapter explores the main theories connected with the roles of emotions and decisions in human behavior and communication, as well as all of the theories and algorithms within artificial intelligence previously carried out by several researchers in search of the “affective decision” in autonomous agents or machines. This chapter also surveys the subject of emotions in human-machine interaction. It highlights the importance of incorporating emotions in human-machine interfaces, as well as generating autonomous decisions and giving way to the “Emotional Machine” concept that a human could be embodied in a machine by special capabilities. In each of these areas, this chapter has also presented the approaches taken in this research. The final section of this chapter demonstrates how research into the automatic analysis of the emotions through the non-verbal aspects of speech, facial gestures, and bodily poses of the human body is related to a wide range of studies supported by machine learning techniques. These different approaches are featured throughout the dissertation, especially in the definition of decision-making affected by emotional feedback. This raised requirements such as the generation of behavior patterns in forecasting the actions of an agent, emotional connotation shared within machine-human interaction, an emotional feedback framework, and the usefulness of emotional information from individuals facing an agent. These requirements are addressed in Chapter 3 to Chapter 8.
3. Decision-Making Model of an Agent

This chapter will focus on a decision-making model as an important issue for the autonomy in an agent. The autonomous agent will perceive its environment and the actions performed by an individual within a common environment. The aim of the decision-making model is to generate consecutive decisions similar to an active behavior. The issued decisions could be affected by external emotions provided by individuals, paving the way for the short-term decision. This approach has a decision analytic flavor, but includes models forecasting the user’s behavior and its impact over the surrounding environment. The model is based on Adversarial Risk Analysis (ARA), a Bayesian approach to strategic decision-making. One builds a model of one’s opponents, expressing subjective uncertainty about the solution concept that each opponent uses, as well as their utilities, probabilities, and capabilities [132]. The decision will be a choice that maximizes expected utility. ARA avoids the standard and unrealistic game theoretic assumptions of common knowledge, using a nested hierarchy of decision analysis models. From the point of view of supporting our agent, the problem is understood as a decision analytic one (see [263]), but we consider procedures that employ the adversarial structure to forecast the adversary’s actions and the evolution of the environment surrounding both of them. On doing this, the agent should forecast what the other participant thinks about him, thus starting the abovementioned hierarchy. Depending on the level to which the agent climbs up in such a hierarchy, we would talk about a 0-level analysis, 1-level analysis, and so on, borrowing the k-level thinking terminology (see [141], [134], [264]). Our approach has a clear Bayesian game theoretic flavor, as in [138] and [139].

3.1 Basic Framework

We have an environment \( E \) that adopts a state \( e \) within a set \( \mathcal{E} \), in which an agent \( A \) and an opponent \( B \), hereinafter referred to as the user, are related through interactions. \( A \) makes decisions within a finite set \( \mathcal{A} = \{a_1, \ldots, a_m\} \), which possibly includes a *do nothing* action and \( B \) similarly in a set \( \mathcal{B} = \{b_1, \ldots, b_n\} \),
which also includes a *do nothing* action. The agent faces this changing environment since this environment affects its own behavior. $A$ has $q$ sensors that provide a reading vector $s_t = (s^1_t, \ldots, s^q_t)$. Analogously, the agent $A$ senses the environmental state $e$ and the readings related to what the user has done based on $a$, possibly probabilistic, transformation function, see more detail in publications I, II and III. The agent plans and carries out its activities according to the basic loop in Fig. 3.1 open to interventions, see [265]. The road of the loop is a group of stages from the inference of the state $e_t$ and the user’s action $b_t$, engaging in actualization in forecasting models until maximizing expected utility, with all of these processes synchronized by a clock.

![Basic Agent Loop](image)

**Figure 3.1:** Basic Agent Loop

### 3.2 Adversarial Risk Analysis Decision Model

The planning of the agent’s activities occurs within the decision analytic framework (see [266]) and incorporates ARA elements, the preference model, and, finally, the corresponding optimization problem.

#### 3.2.1 Forecasting Models

The agent maintains a forecasting model that suggests with which probabilities the user will act and the environment will react, given the past history of the agent’s actions, the user’s actions, and the evolution of the environment and its action $a_t$. We are interested in computing

$$p(e_t, b_t \mid a_t, (e_{t-1}, a_{t-1}, b_{t-1}), (e_{t-2}, a_{t-2}, b_{t-2})),$$

which describes the dependence of the environment and the user action on the agent action and the past two events. According to the mathematical formulation that is described in detail in publication II, Eq. 3.1 may be decomposed into three models.
The first factor, which we call the \textit{environment model}, is fully under the control of the user
\[ p(e_t \mid b_t, e_{t-1}, e_{t-2}). \]

The second factor, which is described in more detail in publication II and in which is incorporated the ARA principle, will give rise to two models \( M_i \) with \( i \in \{1, 2\} \). In these two models, the agent forecasts how the user will react to its actions, and will use this forecast in its decision-making. The first one, \( M_1 \), refers to the user’s reactions to the agent’s actions named \textit{classical conditioning model}, which we describe through
\[ p(b_t \mid a_t). \]

The second one, \( M_2 \), describes the evolution of the user by himself; we call it the \textit{user model} and describe it through
\[ p(b_t \mid b_{t-1}, b_{t-2}). \]

We view the problem as one of model averaging as in [267] and formulated for two steps ahead as shown in publication II and III.

3.2.2 Preference Model

The agent faces multiple consequences \( e = (c_1, c_2, \ldots, c_l) \). At each instant \( t \), with conditions delimited in II. The consequences are formed as
\[ c_i(a_t, b_t, e_t), \quad i = 1, \ldots, l. \]

Assuming that they are evaluated through a multi-attribute utility function [268], without much loss of generality [269], we have
\[ u(c_1, c_2, \ldots, c_l) = \sum_{i=1}^l w_i u_i(c_i), \]
with \( w_i \geq 0 \), \( \sum_{i=1}^l w_i = 1 \), where \( u_i \) represents the robot’s \( i-th \) component utility function and \( w_i \) represents the corresponding utility weight; this takes into account that the agent’s objectives are ordered hierarchically [270].

3.2.3 Expected Utility

Planning \( (r+1) \) instants, the agent aims at maximizing the predicted expected utility. Assuming utilities to be additive over time, several periods ahead could be solved through dynamic programming as in [271] and two periods ahead in
We assumed one period ahead for all of the simulations made in the two versions of the simulator. In this case, we would aim at solving

$$\max_{a_t \in \mathcal{A}} \psi(a_t) = u(c(a_t, b_t, e_t)) \left[ p(e_t | b_t, e_{t-1}) p(b_t | a_t, b_{t-1}, b_{t-2}) \right] db_t, de_t$$

To make the group of the agent’s actions less predictable, probabilities proportional to predictive expected utilities are assumed to be random components \([272]\).

$$P(a_t) \propto \psi(a_t),$$

### 3.3 Implementation

#### 3.3.1 Basic Elements

The agent has 15 alternatives within the set \(\mathcal{A}\) as presented in Fig. [3.2] of which several options show the interaction with the user, e.g., requesting shutting down, telling jokes, and relating stories, to name just a few. Complain actions are also located within this set, e.g., confronted with a disgusting user’s action, the agent can reply with an alert action (see publication [II] for more detail).
The agent is able to detect several actions from the user under the types “interactive”, “aggressive”, and “affective”; they are shown in Fig. 3.3. Fourteen types of actions are captured from the set $\mathcal{B}$. We will assume that this set is fixed, but it could be extended to be implemented within more complex artificial entities such as a robot or a software agent, which can operate in more complex real or virtual environments. In such environments, additional sensors may be needed to infer more complete and realistic sets of actions and environmental states. Regardless of the specific form the sets may take, it is important to mention the alternative of more sophisticated processors that we may use to extend the limited memory of the entity to be able to forecast further steps forward (see more detail in publication [II]).

![Figure 3.3: User’s actions](image)

The group of actions can be handled by deterministic and probabilistic rules. For example, a deterministic rule is shown in move action; it is triggered by the signal provided by touch and inclination sensors. Others are detected according to probabilistic rules, like those involving voice recognition and processing, e.g., shout is triggered when the user’s voice is too loud. The sketch of how the user’s actions are detected, dividing them into deterministic and probabilistic rules, is shown with detail in publication [II].
3.3.2 Forecasting Models

The basic forecasting models (environment, user, classical conditioning) are Markov chains, and we learn about their transition probabilities based on matrix beta models (see [135]). For expected utility computations and point forecasts, we summarize the corresponding row-wise Dirichlet distributions through their means. Learning about probability models is accomplished through Bayesian model averaging.

The Classical Conditioning Model

This model forecasts the user’s actions based on the agent’s action. We will use a matrix-beta model for such purpose [135]. For each \( a_j \), the prior distribution will be Dirichlet with parameters \( \beta_{ij} \geq 0, i \in \{1,\ldots,n\} \), so that

\[
p(b_t \mid a_t = a_j) \sim \text{Dir}(\beta_{1j}, \ldots, \beta_{nj}), \ b_t \in \{b_1, b_2, \ldots, b_n\}.
\] (3.2)

Now, if \( h_{ij} \) designates the number of occurrences of the user doing \( b_i \), when the robot has made \( a_j \) until time \( t \), the posterior distribution will be

\[
p(b_t \mid a_t = a_j, D_t) \sim \text{Dir}(\beta_{1j} + h_{1j}, \ldots, \beta_{nj} + h_{nj}), \ b_t \in \{b_1, b_2, \ldots, b_n\}.
\] (3.3)

The data summary, including its storage and retrieval in a matrix beta model, are detailed in publications [II, III].

The User’s Model

We now provide our forecasting model for the current user’s action based on what the user has done two time steps before. For \( i, j \in \{1,2,\ldots,n\} \), we have a priori

\[
p(b_t \mid b_{t-1} = b_i, b_{t-2} = b_j) \sim \text{Dir}(\rho_{1ij}, \ldots, \rho_{nij}), \ b_t \in \{b_1, b_2, \ldots, b_n\}.
\]

If \( h_{kij} \) designates the number of occurrences that the user did \( b_t = b_k \) after having done \( b_{t-1} = b_i \) and \( b_{t-2} = b_j \), we have that the posterior probability is

\[
p(b_t \mid b_{t-1} = i, b_{t-2} = j, D_t) \sim \text{Dir}(\rho_{1ij} + h_{1ij}, \ldots, \rho_{nij} + h_{nij}), \ b_t \in \{b_1, b_2, \ldots, b_n\},
\]

As well as the previous updating process, the group of data is summarized in a three-dimensional matrix beta model [135]. The process is also detailed in publications [II and III].
Model Averaging

We have two models that can potentially include different basic sets of covariates and different priors; applying Bayesian model averaging [267] we have

\[ p(b_t \mid a_t, b_{t-1}, b_{t-2}, D_t) = \]

\[ = p(M_1 \mid D_t) p(b_t \mid a_t, D_t) + p(M_2 \mid D_t) p(b_t \mid b_{t-1}, b_{t-2}, D_t), \]

with

\[ p(M_i \mid D_t) = \frac{p(D_t \mid M_i) p(M_i)}{\sum_{i=1}^{2} p(D_t \mid M_i) p(M_i)}, i = 1, 2. \]

Under the assumption \( p(M_1) = p(M_2) = \frac{1}{2} \),

\[ p(M_i \mid D_t) = \frac{p(D_t \mid M_i)}{\sum_{i=1}^{2} p(D_t \mid M_i)}, \]

with

\[ p(D_t \mid M_i) = \int p(D_t \mid \theta_i, M_i) p(\theta_i \mid M_i) d\theta_i, \]

where \( \theta_i \) is the vector of parameters of model \( M_i \), \( p(\theta_i \mid M_i) \) is the prior density of \( \theta_i \) under model \( M_i \), and \( p(D_t \mid \theta_i, M_i) \) is the likelihood.

The results of the self-actualization for the two models detailed in publications [II] and [III] are described as

- \( M_1 \).

\[ p(D_t \mid M_1) = p^1_t, \]

We can see that if at iteration \( t + 1 \) the robot performed \( a_j \) and the user performed \( b_i \), the new model probability is updated to

\[ p^1_{t+1} = p^1_t \times \frac{B^t_{(n+1)j}}{B^t_{ij}}, \]

- \( M_2 \).

\[ p(D_t \mid M_2) = p^2_t, \]

then,

\[ p^2_{t+1} = p^2_t \times \frac{\rho^t_{(n+1)jk}}{\rho^t_{ijk}}, \]

assuming that, at iteration \( (t + 1) \), the user performed \( b_k \), after having performed \( b_i \) and \( b_j \).
This process of actualization becomes a fairly stable learning approach by accumulating data. A highly reactive behavior may be obtained by adopting forgetting schemes that focus only on the latest few data observed.

**The Environment Model**

In this case, we assume an agent that carries a group of sensors allowing the estimation of the states that surround it, where the evolution is occurring. The model has seven states, \( e_t = (e^1_t, e^2_t, e^3_t, e^4_t, e^5_t, e^6_t, e^7_t) \); each state of the evolution is the result of combining the data provided by sensors and the rules to which they are subject.

We assume conditional independence for the seven environmental variables, so that

\[
p(e_t | b_t, e_{t-1}, e_{t-2}) = \prod_{i=1}^{7} p(e^i_t | b_t, e^i_{t-1}, e^i_{t-2}).
\]

The states are related with levels of energy, temperature, and variation of light, to mention only a few examples that are addressed with more detail in publications [II] and [III]. For illustrative purposes, we will describe some of them.

- **Temperature** We will assume that \( p(e^2_t | b_t, e^2_{t-1}, e^2_{t-2}) = p(e^2_t | e^2_{t-1}, e^2_{t-2}) \), as we are not able to detect the user’s actions concerning temperature changes. We will assume a simple model of the state, such as \( e^2_t = e^2_{t-1} + (e^2_{t-1} - e^2_{t-2}) \Delta t \).

- **Variation of Light Intensity**

We will assume the generic model of the state \( p(e^7_t | b_t, e^7_{t-1}) \), being \( b_t = blind \), the relevant user action. The light sensor detects whether (1) or not (0) there is sufficient light in the environment. Depending on the detection of light, and whether the agent inferred the user’s action to be blind or another, the next value \( (e^7_t) \) of the light sensor will be predicted according to the rules in Table 3.1.

<table>
<thead>
<tr>
<th>( e^7_{t-1} )</th>
<th>Blind</th>
<th>Not Blind</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 3.1:** Evolution of being lightened
3.3.3 Multiobjective Preference Model

The agent aims at satisfying five objectives, as shown in Fig. 3.4, which, as in [270], are ordered hierarchically by importance. The needs are related to the level of charging, security, acknowledgment, concern, and updating. The agent will invest most resources in achieving a sufficient level in the lowest objective, because of its higher weight. Once it has attained a sufficient value at that level, it will redistribute its resources to achieve the next level, and so on.

![Pyramid of objectives](image)

**Figure 3.4:** Pyramid of objectives.

Utility Function

The objectives described in Section 3.3.3 are formalized in an objectives tree that is detailed in publication [II] which includes the attributes coming from the sensors that the agent has used to access them, e.g., the second objective, concerning security, takes into account whether the noise, temperature, and light levels are appropriate and whether the robot is not being attacked. Therefore, the relevant sensors are the microphone to capture the noise level, the thermometer to measure the temperature of the environment, the light sensor, and the inclination sensor, to verify whether the agent is been attacked or not.
Based on these five objectives, the global utility function would be

\[ w_1 \times u_1(\text{energy}) + w_2 \times u_2(\text{security}) + w_3 \times u_3(\text{be taken into account}) + w_4 \times u_4(\text{being accepted}) + w_5 \times u_5(\text{being updated}), \]

with \( w_1 >> w_2 >> w_3 >> w_4 >> w_5 > 0 \) and \( w_1 + w_2 + w_3 + w_4 + w_5 = 1 \), to emphasize the hierarchical nature of the objectives.

**Component Utility Functions**

For illustrative purposes, some examples of the agent’s component utility functions related to the bottom and the top of the pyramid are represented. More detail about the formulations are described in publications II and III.

- **Objective 1: Energy**

Energy maintenance is a key requirement for an agent to continue operating in a normal interaction with a user. It is a fundamental key to the design of intelligent autonomous systems. The most basic objective pays attention only to the energy level, \( e^1 \), measured on the scale \([0,1]\). A negative indicator is related to a low energy level, while a high energy level is a positive indicator to the agent. Energy is defined as

\[
u_1(e^1) = \begin{cases} 0, & \text{if } e^1 \leq lth \\ 0.5, & \text{if } e^1 = uth \\ e^1 - lth, & \text{otherwise,} \end{cases}\]

with \( uth = 0.5 \) and \( lth = 0.1 \).

- **Objective 5: Being Updated**

With respect to a physical agent as a robot, being updated means that it will check to see if there are any software updates available. The current implementation of such component utility function is

\[
u_5(\text{being updated}) = \begin{cases} 1, & \text{if robot version date} < 2 \text{ months ago} \\ 0, & \text{otherwise}. \end{cases}\]
3.3.4 Results and Discussion

The purpose of the first simulator was to investigate the evolution of the models, assuming that there are some sensors to infer the data provided by the agent-human interaction (see Fig. 3.5). This first version aimed to determine whether the model could cope with a few of the user and agent actions. The simulator limited its scope to five user actions (attack, move, recharge, stroke, and do nothing) and six agent actions (alert, cry, ask for charge, salute, warn, and do nothing). Fig. 3.6, Fig. 3.7, and Fig. 3.8 depict some graphs of the data obtained from the simulator. The graphs are divided into six different charts; in each one, the first couple are 60-minute simulations (A, B), the second couple are 30-minute simulations (C, D), and the last couple are the average of six 5-minute simulations (E, F). As expected, if the user interacts with the agent during the entire time of the simulation as shown in Fig. 3.6, the agent considers the user as very reactive to its actions, so rapidly the value of \( p(M_2) \) achieves the maximum, and vice versa. Regarding the last couple of charts (C, D), we can observe a curious case that could be explained if we look closely at the first iterations close to the value 0, where the values of these two probabilities oscillate until they finally become stabilized. In the case of no interaction as shown in Fig. 3.7, we have obtained unexpected results. The probabilities of both models remain a symmetric behaviour (A, B, C, D). Without any interaction, the behaviour remains constantly anchored in the same state. In the third experiment related to Fig. 3.8, the results are very similar to the first one as in Fig. 3.7 where the agent considers the user as very reactive to its actions. In this case, the average of the six 5-minute simulations (E, F) behaves according to what was expected. Fig. 3.9 shows the expected utility of the different simulations; they vary depending on the interaction of the user. If the user constantly interacts with the agent, its expected utility decreases after 500 iterations, which is possibly due to the mean behavior of the user. If there is an intermittent interaction, its expected utility oscillates randomly. Finally, when there is no interaction, the expected utility of the agent increases exponentially. The model is implemented in an asynchronous mode. Sensors are read at fixed times, with different timings for different sensors. When relevant events are detected, the basic information processing and decision-making loops are activated, as described in Fig. 3.1. The exceptions to standard behavior from the model are managed by the loop being open to interventions through various threads. Taking into account the processing and memory capabilities of the agent, and the need to have almost instant responses, we plan only one step ahead and choose the action with probabilities proportional to the computed expected utilities, to introduce some variety. Memory is limited to the two previous instants.
3.4 Summary

This chapter has described a behavioral model of an autonomous agent that imperfectly processes information from its sensors about the world around it, facing an intelligent adversary (the user/individual) using multi-attribute decision analysis at its core, complemented by forecasting models of the adversary. Improving the user’s experience when interacting with a machine [273] or [123] was our motivation for this model. The model, which may serve as a cognitive personal assistant, may be used for therapeutic purposes and with elderly people for companion purposes. Toward this goal, the agent’s decision-making model requires an emotional input provided by the individual to make its decisions at an emotional level. Recently, the field of cognitive processes has shown that emotions may have a direct impact on decision-making processes; see, e.g., [50] and [274]. Advances in areas such as affective decision-making [6], neuroeconomics [275], and affective computing [2] are based on this principle. Following this, the following chapters will work toward developing an
emotional feedback model to provide emotional factors as a result of the interaction between the agent and the individual. This goal will pursue a more fluent and natural interaction.
Figure 3.6: Evolution of $p(M_1)$ and $p(M_2)$ with the user constantly interacting with the agent.
Figure 3.7: Evolution of $p(M1)$ and $p(M2)$ without user-agent interaction.
Figure 3.8: Evolution of $p(M_1)$ and $p(M_2)$ with an intermittent user-agent interaction.
Figure 3.9: Evolution of the expected utility: with interaction (Top), a mixed interaction (Middle), and no interaction (Bottom).
4. Emotional Acquisition Through Facial Gestures

An autonomous agent as an emotional architecture needs to establish an emotional link with the individual in a common interrelation environment, and this has furthermore been demonstrated in multiple theories [54] that highlight the importance of emotions in machines, as reviewed in Chapter [2]. Emotions play an extremely important role in human lives, and they are reliable indicators of capacity for human socialization. They determine how we think, how we behave, and how we communicate with others. With this information, it is possible to think that the new future generation of machines must have some skills to understand human emotions and act according to them [53], [276]. Humans have an ever-intensifying relationship with automated computer technology, and it is involved in nearly everything we do day after day. Sooner or later, machines will become more clever, and they will fill a growing number of roles in today’s society. Their influence is directly entering virtually every domain of our lives, e.g., surgical assistants [277], helper robots on the battlefield [278], assisting in classrooms in educational contexts [279], nursing homes [280], and offices [281] to name but a few. But to partially enable machines to understand aspects of human emotions, it is necessary to consider the first stage related to their recognition, taking into account the diversity of sources to capture the emotional information. This chapter is the first attempt within our research to recognize human emotion from a detected human’s face by an agent. The output information obtained from this model will be used in Chapter [5] to modify the behavioral system developed in Chapter [3], thereby generating an emotional decision. The methodology uses a classification technique of information pertaining to six universal basic facial emotions from a new fused image. The new fused image is composed of two blocks integrated by the area of the eyes and mouth, very sensitive areas to changes in human expression and particularly relevant for the decoding of emotional expressions. Finally, we use the merged image as an input to a feed-forward neural network trained by backpropagation. Such analysis of merged images makes it possible to obtain relevant information through the combination of proper data in the same image and reduce the training set time while preserving classification rate.
4.1 Implementation of Facial Emotion Detection Model

The emotion detection starts with the input of a new image that passes through a series of phases to become a merged image prepared for analysis in an artificial neural network trained with images from six different facial expressions of emotion per individual. Finally, the system reports the emotional state related to the face. The loop is shown in Fig. 4.1.

![Figure 4.1: Emotion detection loop](image)

The case of a hardware entity (robot) that operates in a real environment differs significantly from a software agent, in which a part of the information provided from the world that surrounds the agent could be virtual, while the robot faces difficult challenges during the entire training period. In this dissertation as well in this chapter, we believe that it is not always necessary to work with a physical agent such as a robot, because it may be possible to usefully simulate an environment, a group of sensors, human actions, etc. In this case, the agent continually focuses on the emotional facial behavior recognition only from the frontal views of the users that pose facial expressions at an advantageous angle, but without face tracking and detection capabilities that allow side facial feature tracking and recognition. The extracted features basically consist of the complete zone of the eyes and the mouth that are merged into a single new image. The resize of the images is carried out through a Nearest Neighbor Interpolation method, and it is applied to all of the images in the training set. For the object of this study, binarization over the images is also important (see more detail in publication IV). An example of the complete process of the facial image treatment can be seen in Fig. 4.2.

![Figure 4.2: Example of the complete process of facial image treatment](image)

Once the images are already known, the next step is to feed the artificial neural network into a feed-forward architecture to find similar patterns in the data. In this case, the recurrent input to the ANN will be pixel-by-pixel, as illustrated in Fig. 4.3. The classification stage consists of three steps. The first step is composed of an input layer containing 1,200 neurons in the form of pixels coming from an image of 40 x 30 pixels. The image is composed of data from
Figure 4.2: Facial Image Treatment

the part of the forehead and the mouth that have been merged into a new image. See more detail about the structure in publication IV.

Figure 4.3: Backpropagation Neural Network architecture

The input variables $x_i$ (each pixel) is multiplied by a weight $w_j$ and are subsequently added. The bias neuron will be the first input $x_1$ equal to 1. The input to each hidden neuron is $z_j$. Refer to publication IV for more detail about the input layer.

The formula can be written as a matrix equation,

$$Z = V^T X \quad (4.1)$$

The second stage is related with the hidden layer and the complexity of its architecture. An approach detailed in publication IV was used to define the
optimum number of hidden layers and hidden units. They depend on the complexity of the network architecture, the number of input and output units, and the number of training samples. The third stage is related to the output layer composed of six nodes as output variables. These nodes are six combinations per individual in a set of facial expressions that are anger, disgust, surprise, happiness, sadness, and fear. Fig. 4.4 depicts how “anger” is processed in six different individuals who belong to the training data set.

![Facial expressions for each individual](image_url)

**Figure 4.4:** Facial expressions for each individual

The output of the neural network is computed as follows:

\[ y_k = \sum_{j=1}^{N_h} w_{jk} r_k \] (4.2)

The formula can be written as a matrix equation,

\[ Y = W^T R \] (4.3)

Now we need to train the ANN by backpropagation, and this involves three stages: the feed-forward of the input training pattern, the calculation and backpropagation of the associated error, and the adjustment of the weights. The data are fed forward from the input layer, through the hidden layer, to the output layer without feedback. The weights \(v_{ij}\) and \(w_{jk}\) of the ANN are initialized randomly, the input \(x\) is taken, and the next step is to find the resulting output \(y\). The desired output \(d\) serves to calculate the backpropagation of associated error \(e = d - y\). The goal now is to minimize the cost function,

\[ J = \frac{1}{2} \sum_{l} e_l^2 \] (4.4)

We assume that the output layer has a linear activation function. Now, we need to adjust the weights of the output layer using the updating rule,
\[ w_{jk}(n+1) = w_{jk}(n) - \alpha(n) \frac{\partial J}{\partial w_{jk}} \quad (4.5) \]

In this equation, the learning rate is \( \alpha(n) \), and affects the speed at which the ANN arrives at the minimum. To find the Jacobian \( \frac{\partial J}{\partial w_{jk}} \), we need use the chain rule, that is,

\[ \frac{\partial J}{\partial w_{jk}} = \frac{\partial J}{\partial e_k} \frac{\partial e_k}{\partial y_k} \frac{\partial y_k}{\partial w_{jk}} \quad (4.6) \]

These three partial derivatives are all relatively easy to find. We have,

\[ \frac{\partial J}{\partial e_k} = e_k, \quad \frac{\partial e_k}{\partial y_k} = -1, \quad \frac{\partial y_k}{\partial w_{jk}} = v_{ij} \quad (4.7) \]

The output layer weights are updated through

\[ w_{jk}(n+1) = w_{jk}(n) + \alpha(n)v_{ij}e_k \quad (4.8) \]

In the hidden layer, we apply a similar principle; the Jacobian is obtained through

\[ \frac{\partial J}{\partial v_{ij}} = \sum_k \frac{\partial J}{\partial e_k} \frac{\partial e_k}{\partial y_k} \frac{\partial y_k}{\partial r_j} \frac{\partial r_j}{\partial v_{ij}} \quad (4.9) \]

with,

\[ \frac{\partial r_j}{\partial v_{ij}} = \frac{\partial r_j}{\partial z_j} \frac{\partial z_j}{\partial v_{ij}} \]

\[ \frac{\partial r_j}{\partial z_j} = \sigma'(z_j) \text{ and } \frac{\partial z_j}{\partial v_{ij}} = x_i \]

The data are actualized through the law for hidden neuron weights, so that

\[ v_{ij}(n+1) = v_{ij}(n) + \alpha(n)x_i\sigma'(z_j) \sum_k e_k w_{jk} \quad (4.10) \]

More detail about the error correction and the weights is presented in publication [IV]. The training data set will be composed of six output expressions per individual in a merged image. The structure is composed of 3 layers in order to train the ANN, and the “Bias Neurons” option and “Sigmoid” for the transfer function were specified. Once the weights are found, the ANN is ready to select the emotion and provide the six outputs related to each set of emotions. The flow of the facial emotion detection process is depicted with more detail in publication [IV].
4.2 Results and Discussion

The MATLAB \([282]\) simulation software was used to simulate the classification process in the emotion recognition. The face’s areas to classify were covered mainly by a complete merged image and the isolated areas of the mouth and eye zones separately. All of these experiments were tested with a database developed by a team of researchers at DSV’s decision analysis group. Finally, facial emotion detection was performed using merged images from the Cohn-Kanade (CK) database \([283]\).

The training set was formed by 36 images of 6 individuals; each individual performed 6 different emotional gestures. The test was composed of 72 new images of 12 individuals, which served as inputs to the artificial neural network. The merging process took into account the areas from the eyes and mouth, because of their greatest potential for providing emotional cues \([284]\).

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>75.00</td>
<td>0.00</td>
<td>0.00</td>
<td>16.67</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.00</td>
<td>83.33</td>
<td>8.33</td>
<td>0.00</td>
<td>16.67</td>
<td>0.00</td>
</tr>
<tr>
<td>Fear</td>
<td>0.00</td>
<td>0.00</td>
<td>75.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Happy</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>83.33</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Sad</td>
<td>25.00</td>
<td>16.67</td>
<td>0.00</td>
<td>0.00</td>
<td>83.33</td>
<td>0.00</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.00</td>
<td>0.00</td>
<td>16.67</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Table 4.1: Confusion matrix using the complete merged image

The measures of the quality of classification were displayed in a confusion matrix that records correctly and incorrectly recognized examples for each type of emotion. The matrix showed some misclassifications, most of them probably due to the diverse types of individuals and the overlapping or entangling through the classification process, which is fully understandable if we take into account the variety of the nature of human emotions and the ways of expressing them. This measure is shown in Table 4.1 and it also served to evaluate the strengths and weaknesses of this system. The more remarkable results with the complete merged image are related to the overall performance of 83.3 % and the 75 % accuracy that the diagonal shows. In some situations, humans misunderstand the information decoded in faces; the case of anger could be mistaken as sadness, bearing in mind the importance of the lips as a main indicator \([285]\). The case of disgust showed relation to sadness, given the similarity of mouth shape in both cases, making arches. The influence of perceiving disgusted and
sad faces is often reflected in important brain regions that are involved in attentional and emotional processes \[286\]. In the case of disgust, we highlight some commonalities \[287\], e.g., in both cases, the eyebrows draw in closer to each other, causing confusion. Some emotions are a mix of expressions, e.g., a fairly common reaction is the form of the mouth and the mixture with fear in the eyes and brow that it is sometimes made during the expression of surprise\[171\]. See more detailed results in publication \[IV\].

To test the efficiency of the algorithm, two different tests were performed, which were made separately in the areas of eyes and mouth, as shown in Fig. 4.5.

![Image data from the eyes and mouth separately](image)

**Figure 4.5:** Image data from the eyes and mouth separately

The test set consisted of 72 new images of 12 individuals, and the training set contained 36 images related to the brows, eyes, and the forehead. This isolated experiment has assessed recognition rates, which can be improved with more information to feed the network. The emotion in the isolated eye zone was recognized with more than 66.67 % accuracy. Emotions like anger, sadness, and surprise were affected in the accuracy of recognition, as shown in the confusion matrix in Table 4.2. Compared with the recognition in merged images, the learning process has required more time in order to identify the desired outputs that contrast with the real inputs; for more detail, refer to publication \[IV\].

In the case of the mouth region, the average successful emotion recognition was 68 %, with more than 58.33 % accuracy. Emotions like anger, disgust,
Table 4.2: Confusion matrix using the isolated zone of eye

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>66.67</td>
<td>0.00</td>
<td>0.00</td>
<td>16.67</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.00</td>
<td>91.67</td>
<td>8.33</td>
<td>0.00</td>
<td>16.67</td>
<td>0.00</td>
</tr>
<tr>
<td>Fear</td>
<td>0.00</td>
<td>0.00</td>
<td>91.67</td>
<td>0.00</td>
<td>8.33</td>
<td>0.00</td>
</tr>
<tr>
<td>Happy</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>83.33</td>
<td>0.00</td>
<td>25.00</td>
</tr>
<tr>
<td>Sad</td>
<td>33.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>75.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.00</td>
<td>8.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>75.00</td>
</tr>
</tbody>
</table>

fear, happiness, and surprise are affected in the accuracy percentage of emotion recognition, as can be seen in Fig. 4.6. The neural network has taken a longer time to train, which was reflected in a higher value of iterations necessary to achieve a greater magnitude of change in the weights. That is comparable in the smaller number of iterations used in the complete merged image. Increasing the number of iterations was favorable to reduce error only slightly and to improve the classification.

Figure 4.6: Percentage of successful emotion recognition in mouth zone

The third test of emotion recognition was carried out using test images from the Cohn-Kanade (CK) database [283]. This database contains 97 individuals showing different expressions. A group of 40 individuals was selected to build the set of 240 images. The algorithm achieved emotion recognition performance of 84%. The percentage of successful emotion recognition is high in emotions like surprise, disgust, happiness, and sadness; this group of results is very similar to the first experiment. The backpropagation error showed a fast
convergence without a high number of iterations. For more detail about the matrix measure and the error depicted, refer to publication IV.

4.3 Summary

This chapter addressed the problem of emotion recognition through facial expression using a method to detect six basic facial emotional expressions by employing an artificial neural network (ANN) in images treated with a merging technique. The recognized emotion will impact the decision-making model of the agent developed in Chapter 3 simulating the emotional behavior in users’ faces when the individual confronts the agent. The image undergoes a processing stage to obtain the features for emotion detection. The features are extracted in the form of a set of pixels to feed the ANN. The use of merged images promotes the selection of relevant information only in the implied areas affected by facial changes. The original image of the face is represented by another reduced-size image that retains the basic features where facial expression shows more changes. These advantages in merging images are shown in a reduced training time, improved performance in classification rate, and fast convergence of the backpropagation error. This method not only allows for further reduction of the size of the image input to the neural networks, but it only considers information that is useful for emotion recognition, discarding unnecessary information that could slow the learning process and negatively affect the recognition performance of the neural network.
5. Decision-Making Model
Affected by Emotional Feedback

Supporting our vision, the autonomous behavior in an agent affected by the external influence of an individual’s emotions must be mandatory to social human-machine interaction. We call this interchange of emotions *emotional feedback*. The agent constantly captures information about the world around it. It makes forecasts about its next decision supported in the Adversarial Risk Analysis (ARA) decision-making model developed in Chapter [3](#). Knowing this and toward the main idea concerning the assistance of the elderly, the agent needs to predict what the user will do next and be ready to provide the best possible interaction, taking into account emotional factors. The conjunction of the emotional feedback provided by the individual’s face and the behavioral decision model of the agent will generate an emotional decision. In light of this remark, this chapter shows an agent able to make decisions influenced by emotional information conveyed through the human face and supported by neural network-based methods in facial emotion recognition developed in Chapter [4](#). The emotional feedback flow will affect the agent’s decision-making and its interaction capabilities with the hint of affective external elements being capable of forecasting the individual’s needs.

5.1 Scenario Focused on the Interaction

Fig. [5.1](#) shows the scenario of the interaction between the agent and a human. Here, the significance lies in the constant flow of information that the agent gathers, which is related to the information of the adversary and the information of the environment in which the agent evolves. The next action carried out by the agent will be emotionally influenced.

In this part of the model, the need arises to create certain emotional influence over the decision generated by an agent. This is performed solely for purposes of creating an emotional connotation effect, which is to say an emotional influence toward a simple action performed by the agent. The agent constantly
stores the information provided by the elderly person, hereinafter referred to as the opponent, to support its final decision. The decryption from human emotional indicators will provide the emotional feedback. For the purpose of this experimentation stage, the methodology used in Chapter 4 was applied as the emotional feedback trigger. The classifier was trained with a set of images from the Cohn-Kanade database composed of anger, disgust, surprise, happiness, sadness, and fear. Each emotion was represented by each person in the training set. The simulation is implemented with four surrounding states that the agent must take into account (Energy, Temperature, Position, and Detection) as described in Chapter 3.

According to Section 3.2 in Chapter 3, \( p(M_i) \) designates the probability that the agent follows the model \( i \), with \( p(M_1) + p(M_2) = 1 \), \( p(M_i) \geq 0 \). From this point onward, the remaining simulations were focused on showing the effect of the emotional feedback over the agent’s reactive behavior facing the opponent’s actions. Utilities and probabilities provided by the agent were also evaluated in terms of the Maximum Expected Utility (MEU) criterion [248], [288] to select the optimal action based on all current information. For more detail about the model, refer to publications III and IV.

The decryption of emotional information brings an added value with the advantage to generate an evolving cycle with a positive behavior when the agent is facing the opponent in an implicit communication, e.g., in the case of a robot caregiver as an intelligent agent, its positive actions can influence the patients that in turn would be more inclined to engage in a closed-loop human-agent emotional interaction [289]. The idea is that the constant adaptation in which the agent could modify its actions appropriately reflects some degree of emotional intelligence [290], [291]. The evolving cycle would be augmented with a certain degree of emotional intelligence that would be implicit within the
agent’s emotional understanding [292], [123], [293], [294], [6].

5.1.1 Rule-Based Design for the Emotional Feedback

The evolution of the agent’s positive behavior is motivated by a self-regulation between actions and emotions that it is supported by broaden and build theory, [295], [296]. In this case, the agent’s reactions when it is facing the opponent’s positive behavior could trigger a cycle of positive emotions connected to the relevant action. Through positive reinforcement, humans are capable of improving skills and capacities in an interrelation, e.g., when humans face challenging situations and they are more likely to make more assertive decisions that when they have a negative emotional charge [28], [297], [298]. As opposed to negative emotions, positive emotions help the body to move from a narrow set of actions to a broader one, allowing the pursuit of a wider array of thoughts and alternatives.

Table 5.1 shows six basic emotions [299], [89] represented in a set of rule-based action/response pairs. The rule describes two groups of emotions organized in a categorical manner, with innate categories found in humans [300], [301]. This table shows the possibility of deciding the agent’s behavior by the effect of emotional feedback over the outputs of the decision system supported by a rule-based algorithm. The pertaining influence over the decisional output would be as a corrective feedback that would create a decision with emotional connotation. Positive emotions may influence the generation of resources and flexible thoughts in order to select the correct action [302]. For more detail about the algorithm and the hierarchy of levels on which the emotional feedback is based, refer to publication [7].

<table>
<thead>
<tr>
<th></th>
<th>$OS_{em}$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$NR_1$</th>
<th>$NR_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>(-)</td>
<td>(+)</td>
<td>(-)</td>
<td>$R_{upL}$</td>
<td>$R_{upL}$</td>
</tr>
<tr>
<td>Sadness</td>
<td>(-)</td>
<td>(+)</td>
<td>(-)</td>
<td>$R_{upL}$</td>
<td>$R_{upL}$</td>
</tr>
<tr>
<td>Fear</td>
<td>(-)</td>
<td>(+)</td>
<td>(-)</td>
<td>$R_{upL}$</td>
<td>$R_{upL}$</td>
</tr>
<tr>
<td>Disgust</td>
<td>(-)</td>
<td>(+)</td>
<td>(-)</td>
<td>$R_{upL}$</td>
<td>$R_{upL}$</td>
</tr>
<tr>
<td>Happiness</td>
<td>(+)</td>
<td>(-)</td>
<td>(+)</td>
<td>$R_{upL}$</td>
<td>$M_L$</td>
</tr>
<tr>
<td>Surprised</td>
<td>(+)</td>
<td>(-)</td>
<td>(+)</td>
<td>$R_{upL}$</td>
<td>$M_L$</td>
</tr>
</tbody>
</table>

*Table 5.1: Emotions, Actions, and Modified Actions*
5.1.2 Results and Discussion

The simulations were carried out using the simulator developed in Chapter 3 that established the conditions of human-agent interaction, in which the rule-based design of emotional feedback was embodied. The agent’s actions were classified by sign (alert(−), cry(−), claim for energy(+), salute(+), warn(+), do nothing(+)), and the opponent’s actions were composed of attack, move, recharge, stroke, and do nothing. The system works per iteration, facilitating the capturing of the opponent’s emotion through the emotion detection model. The number of iterations was fixed at 250, which meant that the agent had 250 opportunities to establish contact with the opponent. The classifier that covers the emotional recognition stage to provide the fixed emotion as input useful to the rule-based design was trained with images provided by the Cohn-Kanade (CK) database [283]. The training stage was performed with 50 images and 250 images as the test set. For more detail about the stage of emotion recognition, refer to publication V.

The hierarchy levels of the agent’s action were distributed from positive to negative in order to set the emotional feedback, which were originally: anger(−), sadness(−), fear(−), disgust(−), happiness(+), and surprise(+) along with their positive or negative sign. Fig. 5.2 is the most representative simulation. It shows that the agent perceives the opponent as very reactive to its actions, so the value of \( p(M_1) \) rapidly achieves the maximum and \( p(M_2) \) the minimum. The simulation also shows the influence of the emotional feedback over the agent’s actions, taking into account that the emotional opponent’s emotional evolution starts at a negative level. The negative component in many actions was suppressed; all of this could be summarized in a positive empowerment of the interaction. For more detailed results, refer to publication V.

5.2 Summary

This chapter described the autonomous behavior of the decision-making model developed in Chapter 3 affected by emotional external outputs. The emotional outputs are provided by the emotions recognition model based on human facial gestures, which is the model described earlier in Chapter 4. The agent’s emotional decision is adapted to the emotional information from the individual, using the bridge of emotional feedback to turn the final decision of the agent into an emotion. In this case, the evolution of the agent’s behavior toward the individual’s behavior takes the positive road. This type of evolution takes into account the reactive response between the emotion and the action, and its positive or negative influence. The emotional output will be the trigger.
Figure 5.2: Effects of emotional feedback

of the corrective emotional output assumed by the agent. The positive behavior that the agent expresses can be seen as a sort of a self-regulation between actions and emotions. The agent’s behavioral actions are consistent with the emotional feedback.
6. Emotional Acquisition
Through Behavioral Human Poses

In human-machine interaction, there is a lack of emotional communication that often results in misunderstanding and remoteness in jointly performed tasks, e.g., health-support activities carried out by a caregiver. The agent is just starting to gain social skills that would enable it to take part in a context of socially intelligent systems. This chapter is developed within a scenario of nonverbal behavioral cues often collected by social communication that establish an approach to predict six basic universal emotions collected by responses linked to human body poses, from a computational perspective. The emotional outputs could be another emotional source to take into account for the decision-making model in Chapter 3. The methodology uses a classification technique of information from six images extracted from a group of videos that have been developed using the Microsoft Kinect sensor for Xbox 360. The methodology was based on the compilation of information provided by body language, taking into account the advantageous information about the emotional state of humans, especially when bodily reaction brings about conscious emotional experiences. The group of extracted images is merged into a single image with all of the relevant information using the approach developed in Chapter 4. The recovered image of human body poses will serve as input to the classifiers, making it possible to obtain relevant information through the combination of proper data in the same image. The method for the analysis of emotional behavior is based on direct classification from the sum of pixels in two-dimensional images previously processed.

Human emotions are strongly connected with bodily states; they are not only connected with mental states, and also show connections at a physical level. Body signals allow us to communicate through non-verbal cues as emotional indicators. The bodily indicators are translated into an informative experience that we could interpret according to the situation in which we find ourselves. The understanding of complex emotions and their physical indicators has been carefully studied in order to analyze and track how we behave in social interactions. However, the challenge is the understanding by a machine of all of this emotional information, which will
influence positively the closed-loop human-robot interaction framework. Cognitive agents are shown to be capable of adapting to emotional information from humans. Taking into account emotional information could be critical when deciding how to handle data captured by a cognitive agent, with direct and precise information of our behavior [31], [32], [33], [34], [305]. How to determine the emotions implied in the behavior analysis of bodily poses is an interesting challenge to investigate, as the link between emotion and the dependencies of bodily poses is very strong. The emotional body language could be a different means of expressing the same set of basic universal emotions as with facial expressions and speech [44], [306], [45]. Components of the body, as three spatial dimensions of body poses and their dynamics, could convey, in unison, the intensity of emotions [307], [308]. Improvements in machines’ communicative behavior could indeed allow people to have a greater closeness to them. This chapter attempts to solve the problem of emotion recognition in humans based only on their more common bodily expressions [27]. In this sense, the approach developed in this chapter can be beneficial to the flow of emotional feedback toward the generation of empathic links between the human being and the agent.

6.1 State of the Problem of Affect Recognition via Body

Human emotions are engines to associate the feeling of others with one’s own emotional internal state and to make decisions. In light of this, machines could also effectively determine their own goal-driven behaviors during social interactions [32], [276], [309]. A live-in companion for the elderly must be able to interact autonomously with a patient with some level of understanding of social behavior and then respond accordingly to the situation. Sometimes, the connection to a social behavior is also carried out in bodily information, as shown in Fig. 6.1 which is generally less understood compared with other type of modalities, notably in the case of face [44] and voice [45] inputs. In some cases, the solution could be the use of a multimodal approach and the merging of multiple sources to recognize emotions with a high degree of accuracy [180]. However, the companion is not always able to react in an opportune and reasonable way, because of dysfunctional behavior of their sensors that makes it impossible to recover all emotional information. Facing a failure of its sensory system, an agent must be capable of switching to other emotional sources to acquire knowledge and decide in advance; against this, the emotion expression recognition via the body could be attractive.
6.2 Data and Methods

6.2.1 Data Set

A controlled study was carried out to evaluate whether the methodology proposed recognizes the emotions in human poses. To simulate the vision sense of the companion agent, the Xbox 360 Kinect Sensor was used [310]. The bodily gestures that revealed the six universal emotions were taken from the bodily motions of 44 individuals in 6 different videos, as illustrated in Fig. 6.2. A diverse set of affective pictures were shown to the individuals as triggers of the stimulus (intensity of pleasantness and unpleasantness) [311]. The database, which serves as the agent’s information, was constructed by taking into account the background related to the bodily motion and its implication in matters of emotions, e.g., emotional-kinetic responses from the human body [312], [313], the link with the neural basis [314], body identity perception in autism [315], sets of bodily stimuli [316], and static views of body poses [42], to mention just a few. For more details about the experimental test, refer to publication [VI].

The kinetics of movement related to different poses can provide a level of knowledge about the associated emotion at the time, e.g., Fig. 6.2 shows a group of six images captured in different time spaces. This was represented
with head curves to the left or to the right, the shoulders pointing downwards and neck hanging to the left or right, supporting the head, waist unbalanced to the left and bent on itself, motionless, passive, with the head hanging on the contracted chest. For more detailed description about the bodily emotions per class, refer to publication VI. The classification is based on the approach developed in Chapter 4. The inputs to the classifier are images of 40 x 30 pixels, hereinafter referred to as matrix-knowledge, as shown in Fig. 6.3. The matrix-knowledge is converted to a row vector of 1,200 features composed by pixels of images captured from the video. The vector-knowledge as a whole feature vector is given by Eq. 6.1 for more detail, refer to publication VI.

$$[VN] = [Vec(MK)]^T$$ (6.1)

6.3 Discussion and Results

To choose the best classification for bodily gestures, five different classifiers (Decision Tree (J48), Bayes Net (BN), Naive Bayes (NB), Multilayer Perceptron (MLP), and Support Vector Machine (SVM)) were tested. For this purpose, the machine learning software WEKA [251] was used. The measures used were stated in Subsection 2.4.2 from Chapter 2. These measures have fa-
cilitated the assessment of the classifiers’ performance, which are principally focused on handling two-class problems. Fig. 6.4 shows that SVM with a linear kernel reached an accuracy of 91.6 %. The same case as a recall that it shows in Fig. 6.5, in which emotions like anger and surprise have achieved the best results using an SVM with linear kernel. All of the classifiers tested have not shown significant differences, but the most representative with a lower training time was the SVM with linear kernel; this is particularly important because this also directly impacts on the building of an agent that entails fewer computational limitations. For more detail about the accuracy and recall of the remaining classifiers, refer to publication VI. Several experiments were carried out to construct the matrix-knowledge, which is made up of six images; more detail about the selection of images is available in publication VI. However, from a computational standpoint would imply less data to process for agent memory.

**Table 6.1:** Confusion matrix

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>39</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Disgust</td>
<td>5</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fear</td>
<td>0</td>
<td>1</td>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Happy</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>41</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Sad</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>Surprise</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>38</td>
</tr>
</tbody>
</table>

The confusion matrix represented in Table 6.1 shows that the rate of correct
Figure 6.4: Accuracy

Figure 6.5: Recall
recognitions over all emotions reached a 91.6 %). Some emotions are perfectly classified, while others are simply confused with others such as disgust and anger with which they are most likely to be confused, e.g., some individuals often express anger and disgust in the same situation, and this situation seems to be a mix of emotions with fear. If the individuals express any emotion mixed with sadness, it is most probable that the result is disgust, anger, or fear [317]. The body expressions in anger and disgust show similarities [299], e.g., the arms are contracted near to the chest. In the case of anger, the position is related to a tentative fight with closed fists, whereas in disgust, the arms are protruded, showing a pushing action; these similarities cause a risk of misclassification between disgust and anger. The same phenomenon is observed between happiness and surprise [318]; studies have demonstrated three cases of overlap, due to some similarities in the upright position from the body and the arms being raised to the sides with forearms straight.

6.4 Summary

This chapter states the relationship between dynamic postures and attributions of emotion in an attempt to describe how emotion may be communicated through the body and usefully collected by an intelligent agent. Part of the image treatment was based on a similar approach developed in Chapter 4 and it does not imply a huge memory consumption in a caregiver agent. The flow of information provided to the agent was based on the similarities between human posture and emotions; this information was considered due to the channel of bodily expression that could be more powerful than other channels of nonverbal communication. Additionally, in this chapter, several classification methods were tested to detect six basic bodily expressions in images provided by a group of videos. The output of the classification stage is an emotion pertaining to a specific body posture, which is another emotional source that could serve as an emotional variable to modify the agent’s decision-making, the cause of the decision-making model developed in Chapter 5. The performance of classifiers was compared in order to select the best result to predict the emotion. The results show that SVM with linear kernel outperforms all of the remaining classifiers, achieving 91.6 % accuracy and a range between 86.4 and 97.7 percent recall. The experiments showed that the classification stage of six images can be sufficient to predict an emotion, in addition to the achievement of high performance and memory saving. Taking into account that the model could be applied to a robotic platform that could make decisions with sufficient computational power and sensing, we suggest the use of SVM because of its flexibility, computational efficiency, and capacity to handle high-dimensional data.
7. Emotional Acquisition 
Through Speech Emotion Recognition

For most humans, nonverbal cues such as pitch, loudness, spectrum, and speech rate are efficient carriers of emotions; these types of features included within the sound of a spoken voice from a speaker could also help a machine to use such properties in recognizing human emotions. This chapter evaluates an approach to decrypt emotions from humans’ speech cues in order to explore another emotional source to feed the decision-making model of an agent developed in Chapter 3. This new emotional source will also support the generation of an emotional decision through the emotional feedback carried out in Chapter 5. The final model later developed in Chapter 8 is based on a mixture of two different sources of external emotional cues of a human being with the internal flow of agent’s decisions, toward the building of an affective companion.

The approach evaluated six different kinds of classifiers to predict six basic universal emotions from non-verbal features located in digital audio files from the eNTERFACE’05 Audio-Visual Emotion Database. The experimental stage was carried out with the aim of selecting the most significant speech emotional features and reducing them to a small group that will serve to build one of the basis classifiers in Chapter 8.

7.1 State of the Problem in Speech Emotion Recognition

Emotional indicators that are embodied in speech may show the emotional state. Once such indicators are learned, theoretically, one can calculate the emotional content of speech that does not depend on speakers or the lexical contents. The decryption of the human voice signal has been the central issue in several investigations, e.g., mental and physiological changes are also reflected in uttered speech [56], [57], [58], psychology and psycholinguistics address frequencies and the intensity of the voice, which show variability of levels across different speakers [206], and short-term spectral features and sound quality could reveal emotional indicators [209], [208].
Recognition of emotions from human speech is increasingly attracting attention and has become an important issue within the companion robotics domain, since human speech provides emotional cues that robots can handle and learn from in social situations; thus, their decisions will be emotionally affected. A companion has to be spontaneous, polite, and must learn how to react according to the human’s emotional charge, providing a friendly environment. Without emotional feedback, it might be challenging for an agent companion to interact with humans in a natural way [319].

To be able to decrypt the emotional information provided by the voice, there exist techniques and approaches for automatic emotional speech recognition that aim to provide the required information to machines and expand their possibilities of interaction. To manage the emotional decryption, the machine learning framework shows several classifiers used in several tasks related to emotion recognition. Each classifier has advantages and disadvantages in order to deal with the speech emotion recognition problem. The more common group used is composed of Hidden Markov Model (HMM) [320], [321], regarded as the simplest form of dynamic Bayesian network, Gaussian Mixture Models (GMM) [322], Nearest-Neighbour classifiers [323], Artificial Neural Networks (ANN) [324], Support Vector Machines (SVM) [325], k-NN [326], Decision Trees, [327] and many others.

7.2 Data and Methods

The emotional speech characteristics were extracted from the eNTERFACE’05 audio-visual emotion database; for more detail, see [328]. The database is based on six universal emotions [329] captured on five different sentences (issued by the individuals) portrayed in 1,320 videos. For this research, only one sentence per each emotion was used, which leads to a total of 264 videos. Each video was subsequently converted to a Waveform Audio File Format using the MultimediaFileReader object from the DSP System Toolbox Library of MATLAB [330]. Fig. 7.1 shows the building process of the data set. For more detail, refer to publication VII.

The data were acquired directly from the group of Waveform Audio files, and they were transformed into 264 vectors of features. The approach that has been designed for a companion agent must contribute to a gradual gain of controllability and robustness that might save substantial cost in computational
efficiency. Using a huge number of features is not a guarantee to achieve the best performance; for that reason, the agent must localize the best features and reject the useless features from the database.

The kind of extracted features used in this research have been commonly used in Music Information Retrieval (MIR). Much of the research is based on the extraction mechanism of features from musical pieces such as timbre, tonality, rhythm, or form, among others, e.g., Spectral Flux (SF) feature \[331\], Spectral Centroid (SC) \[332\], and Spectral Roll off Point \[333\], to mention just a few. For more detail about the features in music information retrieval used, see publication \textbf{VII}. It is attainable that the variability of emotions can be explained by a small set of acoustic features. For this task, in order to identify objective acoustic features, MATLAB was used; most of the employed methods were developed in \[334\]. The final vector of features contains 276 attributes that will be assessed through the classifiers.

### 7.3 Discussion and Results

Six classifiers were tested in a 10-fold crossvalidation scheme. A Support Vector Machine (SVM) has used three kernels, linear and polynomial (with degrees 2 and 3), k Nearest Neighbors (kNN) has used \(k\) from 1 to 15 (showed the best result with \(k=5\)), Multilayer Perceptron (MLP) with hidden neurons from 2 to 20 (showed the best result with 10 neurons), bayesNet (BN), and NaiveBayes (NB) and Decision Tree (J48), as stated in Subsection 2.4.1 from Chapter 2. The several tests with classifiers were conducted using the WEKA.
toolbox. Most of the measures presented in 2.4.2 from Chapter 2 were used to evaluate the performance of classifiers.

The best performance was achieved with the Decision Tree (J48), reaching a 96.21% accuracy facing the other classifiers. (J48) also achieved the most relevant results per emotion positively classified (recall). The features of the tree that were selected, taking into account the information gain, are based on the 2D Method of Moments (2DMM) and the 2D Method of Moments of MFCCs (2DMM – MFCCs). To assess the accuracy of image classification of the decision tree, (J48) used a confusion matrix depicted in Table 7.1 clearly showing a balanced distribution of misclassification rates in the group of emotions, e.g., “happiness” did not show any misclassification, with 100% of emotion recognition. For more detail about the comparison of different classifiers, confusion matrix, and the structure of the tree, refer to publication VII.

**Table 7.1: Confusion Matrix J48**

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happiness</th>
<th>Sad</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Disgust</td>
<td>0</td>
<td>42</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fear</td>
<td>0</td>
<td>2</td>
<td>41</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Happiness</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>44</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sad</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>43</td>
<td>1</td>
</tr>
<tr>
<td>Surprise</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>41</td>
</tr>
</tbody>
</table>

The information gain is visualized as a heuristic to select features, as is done for the decision tree. Taking into account the features selected by the tree, the data set was reconstructed with the selection of the 2D Method of Moments and the 2D Method of Moments of MFCCs. The same classifiers with the same parameters were trained and compared. Also, the best result for each classifier is shown, where the best result for MLP was with 12 neurons and kNN for k=6. Comparisons between the accuracy previously obtained and the results are shown in Fig. 7.2. The accuracies of MLP, BN, and SVM (with polynomial kernel of degree 2) were superior to the Decision Tree. The MLP and BN achieved 96.97% accuracy, and the SVM 96.59%. The recall of different classifiers with all features and with feature selection for each emotion are detailed in publication VII in which the best results are also obtained by MLP, BN, and polynomial SVM with degree 2.
The recall from the four classifiers is shown in Table 7.2, in which the three final rows illustrate the average of recall for each classifier (Average) and the range of the recall (Min and Max). There may not be much difference in accuracy between the four classifiers J48, MLP, SVM with degree 2 (SVM-P2), and BN, but the last three are superior to J48 based on the range of the recall and the average. The range of SVM is equal to MLP, while the average is lower than MLP; this means that MLP has achieved the best results.

Keeping in mind that the building of a system for real time is likely applicable to a companion agent, there is definitely a need to take into account the time consumption of the algorithm. In view of these clear results of the analysis, it would be advisable to use the BN or SVM algorithms instead of MLP to consume fewer computational resources in the future architecture of a companion agent.

7.4 Summary

In this chapter, an approach to performing parameterization of audio data for the purpose of automatic recognition of emotions in speech was developed. The audio database was built with the information of several videos related to human emotional expressions. A group of classifiers was tested to identify
Table 7.2: Comparison of recall for J48, MLP, SVM with degree 2 (SVM-P2) and BN

<table>
<thead>
<tr>
<th></th>
<th>J48</th>
<th>MLP</th>
<th>SVM-P2</th>
<th>BN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>97.70 %</td>
<td>100.00 %</td>
<td>100.00 %</td>
<td>97.70 %</td>
</tr>
<tr>
<td>Disgust</td>
<td>95.50 %</td>
<td>95.50 %</td>
<td>95.50 %</td>
<td>97.70 %</td>
</tr>
<tr>
<td>Fear</td>
<td>93.20 %</td>
<td>95.50 %</td>
<td>95.50 %</td>
<td>95.50 %</td>
</tr>
<tr>
<td>Happiness</td>
<td>100.00</td>
<td>97.70 %</td>
<td>97.70 %</td>
<td>97.70 %</td>
</tr>
<tr>
<td>Sad</td>
<td>97.70 %</td>
<td>97.70 %</td>
<td>95.50 %</td>
<td>97.70 %</td>
</tr>
<tr>
<td>Surprise</td>
<td>93.20 %</td>
<td>95.50 %</td>
<td>95.50 %</td>
<td>95.50 %</td>
</tr>
<tr>
<td>Average</td>
<td>96.22 %</td>
<td>96.98 %</td>
<td>96.62 %</td>
<td>96.97 %</td>
</tr>
<tr>
<td>Min</td>
<td>93.20 %</td>
<td>95.50 %</td>
<td>95.50 %</td>
<td>95.50 %</td>
</tr>
<tr>
<td>Max</td>
<td>100.00 %</td>
<td>100.00 %</td>
<td>100.00 %</td>
<td>97.70 %</td>
</tr>
</tbody>
</table>

the best of this group in emotional prediction. The first experimental stage has covered a feature selection process, in which the outputs from a Decision Tree have been used as a feature selection technique to remove redundant and noisy features. The features provided by the decision tree were 2D Method of Moments and 2D Method of Moments of MFCCs. The feature selection undertaken by the decision tree increased the efficiency of the accuracy and the recall, which enabled the reduction of the dimensionality of the data, leading to fewer computational processes in the agent’s memory. The second experimental stage has been focused on the selection of the classifiers with the best results obtained, which were Multilayer Perceptron, Support Vector Machine, and bayesNet. Support Vector Machine and bayesNet could be good candidates to build the emotional recognition system of an agent, because of their simple implementation and reduced computational complexity. The efficiency of this classification resides in the selection of a small quantity of features from a large group. This could be conceivable for a companion agent that must take fast decisions according to the emotional feedback provided by humans. The agent would face the selection of large quantities of information emanating from facial gestures, kinetics from the human body, and speech signals, or a fusion of these inputs. This classification would work remarkably well on real-world data, facilitating their use in a real-time system.
Humans sense the world through biological senses, each of them with a well-defined specific role. Any information related to a specific physical phenomenon or event that the senses have perceived is routed to the brain entirely commingled, with the possibility to be subsequently decrypted. Sensors can provide the senses for machines; the information provided by them represents vital ways to guarantee the awareness of its environment and to perform its decisions according to an emotional bond with humans. A machine with a significant amount of emotional sources could replace the information that is missing from one particular source. This chapter represents the union of the models and approaches developed in Chapters 3, 4, 5, 6, and 7 in order to probe the viability of the artifact as the entire model proposed. This chapter shows that the artifact should be capable of responding with an emotional decision given the emotional external inputs provided by humans within an Emotional Feedback Framework. The system will track a human’s affective state using facial expressions and speech signals with the purpose of modifying the forecast actions of an autonomous agent. The system uses a fusion of two baseline unimodal classifiers based on bayesNet, giving rise to a multi-classifier. The union of the three classifiers forms a bimodal scheme of emotion classification. The outputs from the baseline unimodal classifiers are combined together through the probability fusion framework applied in the general multi-classifier. The system has been proposed to classify six universal basic emotions using audiovisual data extracted from the eNTERFACE’05 audiovisual emotion database. The relevant emotional information must give to the agent the power to make a decision included in the group of affective decisions.

The search of sources for gathering emotional data in order to decode the emotional states of humans is becoming more extensive, and it is growing in diversity, e.g., the unimodal way shows sources such as speech [335], facial expressions to decode the current emotion [167], [336], the kinetics of postures [337], and physiological brain signals [53], to mention just a few. The diverse
emotional sources can be mixed to enhance their recognition. The fusion of multiple modalities from emotional multi-sensory information can ensure the robustness in emotion recognition processes, similar to those in humans [48]. This is because the data are consolidated in only one source, giving rise to a best classification [234], [338], e.g., in the middle of a foggy environment, blurred vision will be a less useful input than auditory perception.

8.1 System Toward the Overall Problem

The companion agent needs to be capable of inferring the current state of its environment and to know what the patient is doing. These actions must be done within a normal loop of the agent, in which it takes into account three main activities: acquiring process, and deciding using data from the environment and the adversary (patient). To maintain the normal course of the aforementioned activities, the agent needs to be able to interact with the adversary at any time. The emotional connection between the agent companion and the patient is established by the emotional feedback, since the agent is constantly perceiving the emotional state of the patient through speech signals or facial gestures. The agent will behave by taking into account the information of the human adversary and the information of the environment in which the agent evolves. The agent almost constantly forecasts a new action to interact with the adversary. To perform the action, the agent needs to decide the value of the emotional indicators that it has perceived. Fig. 8.1 shows the environment of interaction activities.

8.2 Data and Methods

The audio-visual information was useful to decrypt the emotional cues from the speech signals and facial expressions from the individuals. The intention was to recreate the variable emotional charge within an interaction with humans. The audio-visual data to build the emotional database were collected from the eNTERFACE’05 [328] database; this was the same database used in the experiments of Chapter [7]. Once the set of videos had been selected, we proceeded to acquire the data from the speech and facial gestures separately. Fig. [8.2] shows the process applied to each video to build the audio-visual data set. For more detail about the building of the emotional database, see publications IV, VII and VIII.

All of the images were subject to a merging process developed in Chapter [4] and publication IV, from which the matrix-knowledge was obtained. The matrix-
knowledge is composed of \( n \times m \) real elements such that matrix-knowledge \( \in \mathbb{R}^{n \times m} \); in this case, the real values for \( n \) and \( m \) are \( 40 \times 30 \) pixels, respectively. The matrix-knowledge is converted to a row vector of features, whereby each position is a feature, in which the total amount of features will be 1,200. The matrix-knowledge and the vector-knowledge (VN) are observed in Eq. 8.1 and Eq. 8.2.

\[
MK = \begin{bmatrix}
\sum_{x=1}^{6} Im_{x11} & \cdots & \sum_{x=1}^{6} Im_{x1n} \\
\vdots & \cdots & \vdots \\
\sum_{x=1}^{6} Im_{x1m} & \cdots & \sum_{x=1}^{6} Im_{xmn}
\end{bmatrix}
\]  

(8.1)

\[
VN = [\text{Vec}(MK)]^T
\]  

(8.2)

To recover the information of the speaker’s speech, each video from the group of 264 was converted to a Waveform Audio file format. The tool used to read the video in audio files was the MultimediaFileReader object from the DSP System Toolbox Library of MATLAB [330]. The features extracted are commonly used in Music Information Retrieval (MIR), providing significant amounts of information about the speaker’s mood [339]. MATLAB was used to identify the objective acoustic features, most of them developed in

Figure 8.1: Overall human-agent interaction
The features extracted are: Spectral Flux (SF) feature, Spectral Centroid (SC), 2D Method of Moments of MFCCs, Root Mean Square (RMS), Spectral Centroid Variability (SCV), Zero Crossing rate (ZCR), Compactness, Mel-Frequency Cepstral Coefficients (MFCCs), Method of Moments, Strongest Frequency via FFT Maximum, Spectral Roll-off Point, Strongest Frequency Via Zero Crossings, 2D Method of Moments of MFCCs, Fraction of Low Energy frame, Linear Prediction Cepstral Coefficients (LPCC), and Strongest Frequency Via Spectral Centroid. Some of the features have been subjected to linear transformations to build a final feature vector that contains 276 attributes, which will be evaluated by the classifiers. For more detail about the speech features that are extracted, see [VII].

8.3 Overall System Architecture

We will focus on the activities of emotional support that a companion agent can carry out. The agent must be capable of helping the patient through social interactions, taking into account the emotional information provided by human beings. To accomplish this goal, a three-layer architecture was built, and it is shown in Fig. 8.3. This distribution will allow organizing the system that inte-
grates the decision-making planning of the agent and its emotional acquisition capabilities.

8.3.1 Bimodal System

The agent could be endowed with a sensorimotor system (e.g., with vision, audition, touch sensors, temperature sensor, inclination sensor, and so on) that would allow it to constantly store the information provided by the facial gestures and speech cues from the patients. All of this information is recovered by three classifiers that conform to the bimodal system; two of them are in the unimodal level and the other is a multi-classifier. This first stage is responsible for the recovery of all audio-visual emotional cues important to the final emotional decision at the end of the loop. The stage of unimodal classifiers is responsible for the selection of the audio-visual information cues. To simulate the bimodal classification stage, we conducted several classification tests, using the WEKA toolbox [251]. Taking into account the selection of emotional
cues, the best classifiers were selected in order to be used subsequently as inputs to the multi-classifier. In a real case, an agent can show problems at a particular instant to capture the affective information through its sensors; this could impact the robustness of emotion recognition. To deal with this problem, the system has a decision stage in order to combine its member outputs to feed the next level, the emotional feedback. For more detail about the experimental stage of the multi-classifier, which served to predict six universal emotions from an audiovisual source, refer to publications VII and VIII.

8.3.2 Decision-Making System and Self-Regulation of Emotional Feedback

The right side of Fig. 8.3 shows the interactive decision model. In this stage, the agent may capture information about its surrounding states and the patient (opponent), which it interprets to infer the individual actions and surroundings. As in Chapter 4, the simulation is implemented by taking into account only four surrounding states: Energy, Temperature, Position, and Detection. The models to make the forecasting are opponent’s evolution model and the classical conditioning model based on the ARA framework [340]. The simulations covered the first two hierarchy levels of Energy and Security, as explained in Subsection 3.3.3 from Chapter 3 to construct the Global utility function. To simulate the process, we used the new version of simulator in Python, as shown in Fig. 8.4, in which the agent makes decisions according to the process in the proposed system.

Chapter 5 was focused on the positive evolution of the agent’s behavior, using the main basis of the broaden and build theory [295]. To make more real the interaction between agent and individual, the system needs to have the support of autoregulation (such as decreasing the positive level in an agent’s decision). The effects of broaden and build theory allow to the agent’s decisions a positive evolution, whereas the Emotional self-regulation [341] will be able to achieve in the agent more humanized decisions, as well as the ability to make spontaneous decisions with emotional connotations.

A modification of the rule-based algorithm developed in Chapter 5 is introduced. The new algorithm simply corrects the agent’s behavior using the human biological emotions. Over a considerable period of time, the agent must engage in some form of self-regulation of its forecast decisions. According to the model of emotion-regulation developed in [342], an individual can handle the emotional charge based on the following points: situation selection, situation modification, attentional deployment, cognitive change, and response
modulation. In other words, if the companion agent has noticed that one decision remains stable for a long period, the agent can handle the selection of other decisions randomly from its repertoire of decisions. For the purpose of designing the simulations, we used the algorithm and the repertoire of rules and decisions developed in Chapter 5 and the first two steps of the process model of emotion-regulation, situation selection, and situation modification. For more detail about the Self-regulation of Emotional Feedback algorithm, refer to publication VIII.

8.4 Results and Discussion

8.4.1 Bimodal System

We use the information gain criterion to select the features used by the speech classifier, according to the best results reached by a decision tree in Chapter VII and publication VII. A group of tests was carried out for each case (speech and facial gesture), from which bayesNet has achieved the best results in the two cases. BayesNet was the best candidate to build the baseline unimodal classifiers used by the multi-classifier. Once the multi-classifier had been constructed, we compared the level of accuracy, during which it was concluded that any fusion mode of the multi-classifier was higher than the two baseline
unimodal classifiers. The best classification was achieved by the Probability-Based Product Rule fusion. In Fig. 8.5, the F-measure (the weighted average of the precision and recall) showed that the emotions of disgust and surprise had not reached 100% accuracy; however, precision and recall measures depict that disgust and surprise have achieved the best results. In short, any of the fusion methods (Maximum Likelihood (ML), Probability-Based Product Rule, and weighted average probability) have achieved the best results of the classification through the baseline unimodal classifiers. That is why this method of fusion was used to build the multi-classifier used in the simulations. For more detail about the recall, comparison of accuracy in baseline unimodal classifiers, and precision-recall for each method of fusion, refer to publication VIII.

**Figure 8.5:** F-measure

8.4.2 Decision-Making model Affected by Self-Regulation of Emotional Feedback

Different computational experiments were carried out to show the different behavior caused by the interaction between the patient and the companion agent. The Decision-Making System and Self-Regulation of Emotional Feedback stages worked together in order to produce a decision with emotional connotation. The decision-making system has carried out the task of generating actions that the agent must take. The self-regulation of emotional feedback stage is used to provide the evolution of the agent’s behavior, using the rule-based design of emotional feedback with self-regulation. Each individual has
faced the agent continuously in order, one after the other, showing the six universal emotions ordered simultaneously; this yielded a total of 264 cases where the interaction and the emotional classification took place.

This means that in the future, the agent could learn about the interaction, adapting its behavior to a specific patient. As can be seen in Fig. 8.6, the agent shows negative behavior in actions such as cry and alert prior to the emotional feedback provided by the patients. This will enable the performing of the agent’s actions with negative connotation. The negative component in many actions was suppressed, which is why the interaction was favored in a positive way. For more detail, see publication [VIII]

Fig. 8.7 shows the second simulation that pursues the same objective, but in this case, the agent will face diverse individuals in a random way. Here, the individuals show less interaction with the agent, and this has particularly been reflected between the individuals 22 and 223. After the emotional feedback, a large variety of agent’s actions are allowed. Actions like cry(-) or alert(-) are more permissible when the patient has a positive emotional charge. At the same time, the agent’s actions increase and were transferred to the positive zone; this means that after the emotional feedback, the simulation shows more interaction between the patient and the agent. For more detail, refer to publication [VIII]

Fig. 8.8 shows the third simulation that represents an isolated case from the 44 individuals, in order to see with more detail the self-regulation over the agent’s forecast decisions within the emotional feedback loop. As is reflected in some points of the iteration period, the decisional behavior of the agent changes because of its self-regulation; the agent randomly changed the decision related to its repertoire of decisions. More detail about the self-regulation is available in publication [VIII]

Fig. 8.9 shows the expected utility for this interaction, which is a result of the interaction in this isolated case. Here, we can observe how the expected utility of the simulation varies depending on the interaction with the user. If the user constantly interacts with the agent, its expected utility decreases (interaction points between 0-16 and 123-150), which is possibly due to the mean behavior of the user. The simulation shows, as expected, that the agent perceives the individual as very reactive to its actions. The figure also shows this same finding. The value of $p(M_1)$ (the patient’s own evolution) rapidly achieves the maximum and $p(M_2)$ (the patient’s reactions to the agent’s actions) the minimum for the two cases.
Figure 8.6: Interaction test between the agent and all datasets in order
Figure 8.7: Interaction test between the agent and all datasets randomly
Figure 8.8: Interaction test between the agent and one specific individual

Figure 8.9: Expected utility result of the interaction between the agent and the specific individual
8.5 Summary

The design of emotional interfaces to care for elders is crucial to achieve successful support in cognitive rehabilitation schemes. This chapter proposed a concept of a system for a companion agent that could select the appropriate actions, regarding the emotional cues that emerge from interactions with patients.

The model uses multi-attribute decision analysis, forecasting models of the adversary and emotional feedback provided by the opponent, all supporting the final decision of the companion agent. The proposed emotional acquisition stage consisted of preprocessing, feature extraction, and pattern classification steps. Preprocessing and feature extraction methods were devised so that emotion-specific characteristics could be extracted in an audio-visual scheme.

A bimodal system fuelled by emotional cues provided by facial gestures and speech signals was proposed, following the approaches developed in Chapters 4, 6, and 7.

There is a chance that the companion could operate under certain circumstances in a noisy environment or with the malfunction of a group of its sensors. Against this background, the companion could react based on the emotional information of one or two sources, or the fusion of the two sources. To deal with that situation, a bimodal system composed of two baseline unimodal classifiers and a multi-classifier was proposed.

Six classification methods to classify emotions from speech signals and facial gestures were tested to build the unimodal classifiers, among which bayesNet was selected. The high recall value of unimodal classifiers, a multi-classifier supported by two baseline unimodal classifiers (each one trained per single source), was proposed. The multi-classifier used three methods of output fusion: Maximum Likelihood (ML), Probability-Based Product Rule, and weighted average probability. The Probability-Based Product Rule achieved the best results, with 98% accuracy, and precision ranges between 97-100% and 97.7-100% in recall.

According to the final architecture that has been proposed, the agent is capable of improving its social interaction by changing its behavioral decision state according to the affective state of the individuals that it is facing. The agent’s final decision could be considered as an affective decision, taking into account that the agent predicts the affective consequences of each available al-
ternative to not disturb the individual to which it is facing. This could only be interpreted as strengthening the relations between the human and the machine within an emotional loop.
9. Concluding Remarks

9.1 Concluding Summary

We assume that the agent is situated within an environment in which there is an individual with whom it interacts, and that it is endowed with a sensori-motor system through which it may capture information about its environment and the individual that it is facing. It interprets the information to infer the actions and emotions of the individual as well as the state of the environment. We are able to come to the conclusion that information, on one hand, could be evaluated to discern its impact on the agent’s objectives, learn the evolution of the environment, and make forecasts affected by the emotional behavior of the individual. Forecasts and evaluations could be affected by external emotions, and external emotions and forecasts (and the agent value system) have an impact on its decision-making process toward building a decision-making emotional system for the agent. We can conceive of creating models capable of learning from the emotional behavior of humans and proceeding according to this behavior. The approach aims at combining and expanding recent developments in decision-making and emotional computing to building decision-making models that take into account the emotional feedback from humans. Under the ARA perspective and emotional feedback, we will consider flexible agents that may have a more or less social behavior, within a unifying model. Such variability in agents’ behavior will depend upon and evolve from the emotional state of the individuals, which, in turn, will impact the agent based on its internal decisions.

Artificial neural networks as machine learning classifiers to recognize human emotions based on facial images are useful for practical implementations in which the dimensionality of the picture is reduced. To reduce the size of the images, a resizing pixel technique was proposed in which the original large-size facial image is represented by another reduced-size image, retaining the basic features of the face so that the person could even be recognized by this image. This technique is different from other methods based on face descriptors, where the face is represented by a set of coefficients from which the face can be reconstructed. Also, considering that human emotions are mostly rep-
resented through the eyes and the mouth, only these portions of the face have been considered in the emotion recognition process. Other parts of the face containing unnecessary information have been eliminated. This method not only allows for further reduction of the size of the image input to the neural networks, but it only considers information useful for emotion recognition, discarding unnecessary information that could slow the learning process and negatively affect the recognition performance of the neural network, which could have a direct impact on the agent’s memory. The first attempt to decrypt emotions in Chapter 4 used neural networks as static networks where there is not a temporal relationship between inputs and outputs.

The features used for emotion recognition in Chapters 6 and 8 to describe each emotional case are based on a matrix as the result of combining the pixel matrices of different images. We show that six images can be sufficient to predict emotion, achieving a high performance and saving memory, and considering an inexpensive robotic platform. Taking into account that the model could be applied to a low-cost robotic platform that could make decisions with sufficient computational power and sensing, we suggest the use of SVM because of its flexibility, computational efficiency, and capacity to handle high-dimensional data. The cost directly affects the technology acceptability; thus, innovation by using less expensive computer systems, sensors, and computational capabilities is relevant and should be taken into account in the implementation of robotic systems.

Even though the emotional outputs from the experimental findings carried out in Chapter 6 have not been used in the fusion of the emotional sources towards the Decision System and Self-Regulation of Emotional Feedback, we can conclude that the results of the experimentation in Chapter 6 have clearly shown links between nonverbal bodily behavior and the emotional content. We have shown that the emotional triggers systematically affect the patterning of human body poses. The information conveyed by the body modality contains large amounts of emotional data, compared with what has been assumed until now. The results suggest that there may be few emotion-specific prototypical patterns of body postures in humans that are clearly visible and identifiable. Human-based body movements and postures could represent the social interpretation of emotions that would be effectively used by an agent.

The output information obtained from the ensemble system could feed an input to a machine capable of interacting with forecasted social actions, in the context of building socially intelligent systems capable of conceiving a more intimate connection between people and machines. The learning of the indi-
individual’s behavioral patterns in a real emotional interaction with an agent serves as support to decide the group of agent’s decisions pertinent to the individual. The emotional cues extracted for two or more emotional sources could bring to the agents emotionally intelligent capabilities. Chapter 5 and 8 show the agent’s emotional decisions toward a positive evolution within the interaction with the individual. The analysis of merged images and the features extracted from the speech samples make it possible to obtain relevant information from the emotional evolution. It is shown by experimental results that the proposed system can detect emotions with satisfactory accuracy, achieving the change of the emotional behavior of the agent faced with a human.

The computational complex of the machine learning algorithms analyzed in Chapters 4, 6, 7, and 8 show significant differences between the testing and forecasting processes. The majority of these algorithms show a linear complexity to make forecasts once they have been trained; that is to say, an $O(n)$, except that $n$ is different in each one. In the case of the decision tree, $n$ is related to the size of the tree or the number of generated rules. In the case of $k$NN, $n$ is the number of cases found in the database. For a neural network, $n$ is the number of nodes in the network, similar to the Bayesian network. In Support Vector Machines, $n$ depends on the number of features and support vectors in the case of a linear kernel and only the number of features in the case of linearity. Either way, the linear complexity allows the response to occur in real time with great speed. Furthermore, the proposed ensemble classifiers in Chapter 8 form a linear combination of the baseline classifiers’ outputs; this means that the complexity is also linear. In this research, the computational complexity of training time $O(sn)$ is related to the models used, because it is a self-learning process, e.g., the training process with new cases in the robot’s database could be made during leisure or charging time. In the case of $k$NN, the complexity suffers from a constant growth because of the additional cases in the database.

The system could operate in a dynamic, nondeterministic, real-life environment, capable of responding in a timely manner according to the emotional features of individuals, and with a suitable degree of autonomy to correct gaps in its knowledge, as it is based on a decision-making model that uses an evolving forecasting model affected by emotions. The system is deeply grounded in human emotion studies that complement novel decision theoretic models. We have tackled both hard scientific and technological research issues to develop a possible agent fit for rendering high-quality emotional links, specifically in the field of companion services. The goal of this dissertation was to develop sound approaches and techniques for emotional decryption and to assess their impact in individual decision-making, underpinning the development of robotic sys-
tems to provide pertinent design paradigms.

9.2 Future Research

Continuing with the core of the work presented in this dissertation, associated future work could take several paths, any of them connected with the deep research in autonomous decision-making and emotional self-support systems. One road will take into account the connectivity of smart devices with humans; such interaction will have to be taken into account in the future, because the number of devices connected wirelessly with intelligent machines has been increasing exponentially.

All of the multimodal sensor inputs that recover biophysiological signals from humans can be used with optimized algorithms to embed affective intelligence in synthetic agents. As a case in point, the wireless sensor networks can improve the sensory capabilities in robots. These groups of sensors are akin to standard sensors, but with better monitoring and control of a wide range of information. The OpenFlow connection environment could serve as the link between a more precise sensing of emotional signals from humans and their proper interpretation. With all of these data collected, we can explore more deeply the affective communications and develop emotional companion systems with more emotional capabilities that will connect with humans, offering improved natural interaction.

Handling emotions could enhance the productivity, personal power, and quality of life of human beings. Based on this concept, keeping in mind that if the smart devices could acquire emotional abilities and would be able to manage the control of the criteria of their actions, this would allow them through judgments of value to adapt their behavior to improved situations. Through inexpensive and wearable sensors, the Internet of Things (IoT) can be used to acquire a high level of knowledge with respect to emotions. Data are valuable to the process of sentiment analysis procedures linked with affective computing and extracting the subjective information from diverse smart sources; this could be used in healthcare, telemedicine, or smart well-being systems that are seen increasingly often.

As the main field of the future research, we will consider a smart interconnection platform that could serve as emotional motor knowledge in order to understand autistic children, to describe more precisely their internal states. To incorporate new models to the core of the basic behavior of autistic emotions, each device would evolve emotionally according to such models that collect
the emotional data from the child. These sensors allow us to have direct feed-
back, which permits us to see this internal future state in a very concrete way. 
This type of emotional device will have the potential to reveal more about emo-
tional state and the early detection of crisis, balancing lifestyle and regulating 
stress level.
Sammanfattning

av informationen om känslomässiga tillstånd som ges av klassificeraren vilket ger upphov till “emotionella beslut”.
References


References:


[68] L CHRISTENSEN AND B JOHNSON. Educational research: Quantitative, qualitative, and mixed approaches, 2004. [40]


[71] HC MARAIS. Basic Concepts in the Methodology of the Social Sciences, 4. HSRC Press, 1988. [40]


[77] H RUSSELL BERNARD. Social Research Methods: Qualitative and Quantitative Approaches. SAGE Publications, 2012. [41]


[305] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. Can your smartphone infer your mood. In Proceedings of the PhoneSense Workshop, 2011. 114


