AMP : An Appraisal-Mood-Personality Model of Emotion

[1]Arjun Verma
Project Supervisors : Dr. Moojan Ghafurian and Dr. Kerstin Dautenhahn
Social and Intelligent Robotics Research Laboratory
University of Waterloo

Mitacs Internship Report

Abstract. *This technical report contains a brief overview of the work conducted during the Mitacs Globalink Internship at SIRRL, University of Waterloo. The goal of this project is to make humanoid robots emotionally intelligent by building a computational model of emotion (CME) to be integrated into them. The technical report entails a survey conducted of the existing computational models of emotion. The survey helped in identifying key components and characteristics to be integrated while developing the CME. We then move on to the details of the proposed framework - AMP. The report also talks briefly about the dataset on which the model is being tested currently.*

1. Introduction

Robotics is a technology that has brought great advancements to the world and the way we live. At the dawn of robotics research, great emphasis was put on making these robots useful in industrial settings. This led to the onset of automation and the very first instances of introduction of machine intelligence in the industrial workspace [Chen and Poo 2003]. The recent decade has however seen a rise in popularity of robots being deployed not only in industries but also at homes and other social settings [Pino et al. 2015]. A few examples of these settings include robot-assistive care for dementia patients [Hung et al. 2019] and tutelage of children diagnosed with autism spectrum disorder [Wood et al. 2021]. With the deployment of robots in such sensitive social contexts, the ability of such robots to interact with humans in ways that resemble human interaction becomes increasingly more relevant [Breazeal 2009]. Such applications call for robots to not only be intelligent (to make optimized decisions) but also emotionally intelligent (to make socially appropriate decisions). The ability to portray emotions allows such robots to be placed in the context of the *Affective Loop* to perform better social interactions. The *Affective Loop* is defined by [Höök 2009] as the interactive process in which "the user of the system first expresses her emotions through some physical interaction involving her body, for example, through gestures or manipulations; and the system then responds by generating affective expression, using for example, colours, animations, and haptics" which "in turn affects the user (mind and body) making the user respond and step-by-step feel more and more involved with the system."

Current work on integrating aspects of emotion into social robots can be broadly classified into three categories : *Emotion Recognition*, *Emotion Expression* and *Emotion Generation*. While there have been significant contributions in the domains of *Emotion Recognition* [Spezialetti et al. 2020] and *Emotion Expression* [Paiva et al. 2014], there is a clear lacunae when it comes to *Emotion Generation*. This is evident even in the aforementioned definition of *Affective Loop*. The statement encompasses aspects of emotion recognition and expression but completely bypasses the generation or the synthesis of these emotions.

The lacunae in the research of emotion generation arises not only due to a lack of address but also because of the absence of conformity and acceptability among the varying existing computational model of emotions (CMEs). This lack of conformity is due to the vast inter-disciplinary background required to come up with these models. These models are built upon researchers' own assumptions of underlying psychological theories of emotion. Each of these models come up with their own characterizations and representations by which the inputs are to be perceived and processed. This has resulted in a wide variety of research in vastly differing and conflicting directions [Marsella et al. 2010] instead of a single grail of integrable components to be followed and built upon. In this technical report

2. Background

Over the past two decades, several computational models of emotion have been proposed to enable artifical agents and social robots to generate emotions of their own. However, before we reach the step of comparing and identifying important characteristics from these models, it is necessary to introduce the theories of emotion formation and how they are connected with robotics. A step by step guide to build the necessary background for developing a CME can be obtained by understanding the various : *theories of emotion*, *theoretical emotion classifications*, existing *emotional robots* and finally, existing *computational models of emotion*. We limit the scope of this technical paper to only the relevant architectures and theories for understanding the proposed AMP framework. Therefore, in this section, we only provide a brief tabular summarization for each of these steps. The summarizations have been obtained from detailed survey papers and one can refer to them for more information. The relevant architectures behind the inspiration for the proposed framework have been discussed in the subsequent section.

2.1. Theories of emotion

In order to get plausible explanations behind the emotional aspects being modeled, researchers usually follow particular theories and models of emotion. These theories impart insights into building architectural and functional designs of CMEs by providing explanations behind key aspects of the emotion generation processes in humans. Each of the varying theories explain the process of human emotions from different perspectives and at different levels of abstraction [Rodríguez and Corchado 2013]. However, these theories usually hold contrasting positions in their explanations of these emotional aspects. Thus, it is of utmost importance to understand the theoretical foundations behind contemporary CMEs to recognize their relevance as well as integrability with components of other CMEs to build further advanced models.

The major theories of emotion can be grouped into three main categories [Myers 2004] :

- physiological: responses within the body that cause emotions
- neurological: emotional responses due to brain activity
- cognitive: thoughts that form or influence emotions

Theory	Summary	Example
James–Lange theory (James, 1884)	Emotions occur as a result of physiological reactions to events	(1) You are walking through a parking garage (2) You notice a dark figure and your heart begins to beat faster (3) You interpret your physical reactions as fear
Cannon–Bard theory (Cannon, 1987)	Physical and psychological experience of emotion happen at the same time	(1) You are going on a date with someone you like (2) You feel excited and have sweaty palms
Schachter–Singer theory (Schachter and Singer, 1962)	Physiological arousal occurs then the individual identifies the reason for and accordingly labels an emotion	(1) Your heart is beating faster during an exam (2) You realise this and identify the emotion as anxiety
Cognitive appraisal theory (Smith and Ellsworth, 1985)	Thinking must occur before experiencing an emotion	(1) You encounter a bear in the woods (2) You think that you are in great danger (3) You experience the fear and fight-or-flight reactions
Facial feedback theory (Strack et al., 1988)	Facial expressions are connected to experiencing emotions	(1) You are being forced to smile at a social event (2) You have a better time than when having a stolid face

Figure 1. Five major theories of emotions and their examples [Yan et al. 2021]

Based on these categorisations, a summary of five major theories of emotion is given in Figure 1. To further illustrate the various emotion theories, a flowchart outlining the different theories and their interactions with various stimuli is depicted in Figure 2.

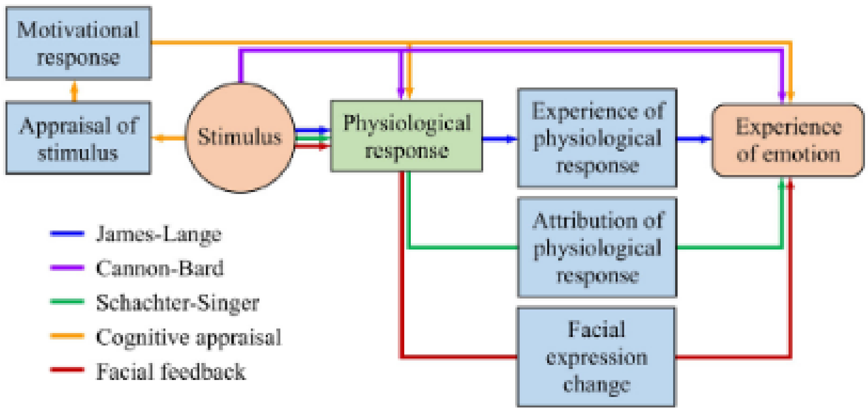

Figure 2. Illustration of the five theories of emotion [Yan et al. 2021]

2.2. Classification of Emotions

After deciding upon an underlying theory of emotion on which the CME would be built, one would need to determine the different categories of emotions that the CME would generate. Different researchers express different views and put forward varying theoretical views on emotion classification [Ortony and Turner 1990]. A tabulated summary of popular theoretical models of classifications of emotions and examples of their applications in robot emotion are presented in Figure 3. The tabulation is presented in a chronological order to show how the growing popularity of artificial emotion intelligence has led to the development of this field which resulted in these classifications becoming more elaborate with time.

Model	Basic emotions	Applications
Arnold (1960)	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness	Belkaid et al. (2019)
Izard (1971)	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise	Djara et al. (2018)
Plutchik (1980)	Anger, anticipation, joy, trust, fear, surprise, sadness, disgust	Mcginn (2020)
Ekman et al. (1982)	Anger, disgust, fear, joy, sadness, surprise	Coco et al. (2018)
Tomkins (1984)	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise	Hwang et al. (2013)
Frijda (1986)	Desire, happiness, interest, surprise, wonder, sorrow	Canamero (2019)
Oatley and Johnsonlaird (1987)	Anger, disgust, anxiety, happiness, sadness	Robert et al. (2018)
Robinson (2008)	Interest, alarm, attraction, aversion, surprise, indifference, hope, fear, gratitude, anger, joy, sorrow, relief, frustration, pride, embarrassment, generosity, avarice, sympathy, cruelty, love, hate	Stebbins (2018)
Cordaro et al. (2018)	Amused, angry, awe, bored, confused, contempt, content, coy, desire, disgust, embarrassed, fear, happiness, interested, pain, pride, relief, sadness, shame, surprise, sympathy, triumph	Cowen et al. (2019)
Cowen and Keltner (2017)	Admiration, adoration, aesthetic appreciation, amusement, anger, anxiety, awe, awkwardness, boredom, calmness, confusion, craving, disgust, empathic pain, entrancement, excitement, fear, horror, interest, joy, nostalgia, relief, romance, sadness, satisfaction, sexual desire, surprise	Eisenberg et al. (2019)

Figure 3. Classifications of emotions and their applications [Yan et al. 2021]

2.3. Emotional Robots

After obtaining the relevant theoretical background on the processes of emotions, it is important to throw some light on existing work of integrating emotions into robots. As per [Mitsunaga et al. 2008], emotional robots must have the capability to recognise human emotions and express their own emotions which results in less mechanical and more natural communication between humans and robots. Furthermore, [Yan et al. 2021] defines the main functions of emotional robots to be as follows :

- Speech recognition: interpretation of human speech and its translation into text or commands for/by robots.
- Verbal interaction: exchange of messages between humans and robots via sounds or words.
- Facial identification: identification of human emotion by analysing patterns based on one's facial attributes, such as textures and shape.
- Emotional expression: exhibition of behaviours that communicate an emotional state or attitude of robots.
- Action coordination: responses that reflect bodily movement of a robot corresponding to the emotional transition.

We again note that while the given definitions address aspects of emotion recognition and expression, the mechanism of synthesis of emotions has been completely overlooked. With the definitions stated, we now present a tabulated summary of existing emotional robots over the past two decades in a chronological order highlighting their functionalities Figure 4.

Figure 4. Emotional Robots and their main functions [Yan et al. 2021]

2.4. Computational model of emotions

In this subsection, we give a tabulated summary of various computational models of emotion Figure 5. The summary includes the base cognitive theory used by the model, emotion theory on which the emotion elicitation and generation mechanism is based on and some of the notable cognitive/behavioural effects modelled in the proposed emotion model. This table just highlights the existing work on computational models of emotions, the relevant architectures for AMP framework have been discussed in detail in the next section.

Model	Base cognitive theory	Emotion theory	Effects modelled
EM [104]	Oz architecture	OCC	Plan change
ACRES/WILL [79,127]	BDI, Planning, Decision Theory, Agents	Frijda	Coping: goal shift, attention shift
Cathexis [130,131]	BDI	Izard, Tomkins, Roseman	Behaviour modulation, habituation and sensitisation, emotional conditioning
Émile [37]	Strips Planning	OCC, Sloman	Plan change, plan selection criteria
FLAME [30]	Fuzzy logic, Planning, Decision Theory, Q-Learning	OCC, Roseman	Choice and inhibititon of plans, emotion-based learning and conditioning
H-CogAff [120]	BDI, Cognition and Affect	Sloman, OCC	Attention shift (alarms), decision biases, precognitive reaction
ALEC [33]	CLARION	Sloman, Damasio	Decision rules learned based on past experience
TABASCO [95]	ACT, BDI	Scherer, Lazarus, Smith	Plan updates
EM-ONE [119]	Minsky-Sloman	Minsky, Sloman	Modification of "narratives": plans, desires, or beliefs. Modifications of "critic" processes
KARO [75]	BDI	OCC, Logic of Emotional Agents (LEA)	Plan/agenda changes; Fear causes cautious planning
MAMID [49–51]	Belief Net, Decision Theory	Scherer, Smith & Kirby, Sloman, OCC	Biases mental constructs (data) based on emotional state; Working memory capacity, speed; attention
ActAffAct [100]	Agents, BDI, Unified Cognition	Frijda, Scherer	Coping: choice of Relational Action Tendency
EMA [70]	BDI, Agents, Decision Theory, Planning	Lazarus, Smith, OCC	Coping: attention shift, plan changes, BDI changes, action tendency changes
Soar-Emote [67]	Soar, PEACTIDM	Scherer, Roseman	Attention shift, goal shift, reinforcement learning biases (both encoding and recall)
WASABI [8]	BDI	PAD, Scherer	Plan utility valuation process biased towards optimism or pessimism, mapping of emotions as beliefs, action biases
TAME [83]	BDI	Unsure	Emotion, mood and behavioural dynamics
FAtiMA [22]	BDI	OCC, Scherer, Lazarus	Coping: plan and goal changes
Schneider and Adamy [117]	Fuzzy logic	Unsure	Motivation based goal achievement
MA/SDEC [114]	BDI, Social contagion	Hatfield, Cacioppo, Rapson	Effect of social contagion in multi-agent interaction scenario
Rasool et al. [101]	Fuzzy logic	Russel, Mehrabian	Expression of empathic responses
EMIA [54,55]	BDI, Fuzzy logic, Gross's Emotion Regulation	OCC, Scherer, Roseman	Emotion transition, emotion regulation
Sun et al. [126]	CLARION	Scherer	Coping, action/behaviour selection
CAAF [56]	Belief-Desire Theory of Emotion (BDTE), Cognitive Agent programming Frameworks(CAF)	Reisenzein	Goal directed behaviour
AfPL [36]	BDI, Probabilistic modal logic	OCC	Goal directed behaviour
GenIA^3 [3]	BDI, Logic	OCC	Goal directed behaviour

Figure 5. Existing Computational Models of Emotion [Ojha et al. 2020]

3. AMP Framework

With a thorough understanding of the relevant background and existing CMEs, we now begin the description of our proposed framework - AMP. In this section, we first begin with the identification of key characteristics and theories to be integrated into our model. We then move onto the inspiration behind implementation of each of these characteristics and how we can leverage designs from existing CMEs. Lastly, we focus on the integration of all these aspects together into our proposed framework.

3.1. Identifying Key Characteristics and Components

3.1.1. Emotion Theory - Appraisal Theory

The appraisal theory is one of the most highly regarded and accepted theory of emotion when it comes to building computational models of emotion (see Figure 6). Appraisal theories explain the elicitation of emotions on the basis of the relationship between individuals and their environment [Ortony et al. 1988], [Frijda et al. 1989], [Roseman et al. 1990]. Appraisal theories suggest that emotions arise from the evaluation of situations, objects, and agents existing in the environment that directly or indirectly impact the individual's goals, plans, and beliefs. This assessment of the individual - environment relationship is carried out using a series of appraisal dimensions such as pleasantness, goal conduciveness, suddenness, controllability, and self-responsibility. Appraisal dimensions can be seen as measurement variables that help to extract and derive information about the influences between the agent and its environment [Frijda et al. 1989]. The appraisal variables may vary in number and type across implementations, however they must be sufficient to collect information about elicitation and differentiation of emotions [Scherer 2001]. We use the appraisal theory as one of the foundational theory of emotion behind our framework.

Model	Theoretical foundations	Appraisal dimensions	Emotions
EMA [38]	Appraisal theory by Smith and Lazarus [44]	Relevance, perspective, desirability, likelihood, expectedness, causal attribution, controllability, and changeability	Surprise, hope, joy, fear, sadness, anger, and guilt
Flame [45]	Appraisal theory by Ortony et al. [12] and Roseman et al. [42]. Inhibition model by Bolles and Fanselow [46]	Desirability, expectation, causal attribution, and standards (or social norms)	Joy, sad, disappointment, relief, hope, fear, pride, shame, reproach, and admiration. *Complex emotions*: Anger (sad + reproach), gratitude (joy + admiration), gratification (joy + pride), and remorse (sad + shame)
Mamid [47]	Appraisal theories such as those by Lazarus [43] and Smith and Kirby [48]. Personality models such as the five factor model [49]	*Universal*: novelty, unexpectedness, intensity, threat level, desirability. *Individual and context-dependent*: individual history, experience, and expectation and goal congruence	Anxiety/fear, anger/aggression, negative affect (sadness, distress), and positive affect (joy, happiness)
Alma [39]	Appraisal model by Ortony et al. [12], five factor model of personality [49], and PAD temperament space by Mehrabian [50]	Desirability of events, praiseworthiness of actions, appealingness of objects, liking of objects, likelihood of an event occurs, and realization that an event has occurred	Admiration, anger, disliking, disappointment, distress, fear, fears confirmed, gloating, gratification, gratitude, happy for, hate, hope, joy, liking, love, pity pride, relief, remorse, reproach, resentment, satisfaction, shame
Cathexis [37]	Diverse psychological [42, 51] and neuropsychological [1] theories	The model does not report used appraisal dimensions but states that they are based on Roseman et al. [42]	*Primary emotions*: anger, fear, sadness/distress, enjoyment/happiness, disgust, and surprise. This model generates *secondary emotions* but does not provide an explicit model for the labeling of them
PEACTIDM [52]	Appraisal theory by Scherer [13] and physiological concepts of feelings by Damasio [1]	Suddenness, predictability, goal relevance, intrinsic pleasantness, outcome probability, causal agent, causal motive, discrepancy from expectation, conduciveness, control, and power	This model implements the model by Scherer [13] for the mapping of appraisal dimension values to specific model emotions
WASABI [36]	Appraisal theory by Scherer [13], PAD space by Mehrabian [50], and physiological concepts by Damasio [1]	Intrinsic pleasantness, expectation deviation, goal conduciveness, dominance, causal (mis) attribution, and coping	*Primary emotions*: angry, annoyed, bored, concentrated, depressed, fearful, happy, sad, surprised. *Secondary emotions*: hope, fears confirmed, relief

Figure 6. CMEs inspired by appraisal theories [Rodríguez and Corchado 2013]

3.1.2. Dimensional Theory for Emotion Classification - Mehrabian's PAD model

The second foundational theory that we integrate into our framework is for emotion classification and description. We use the three-dimensional framework proposed by Russell and Mehrabian [Russell and Mehrabian 1977]. This model describes emotions based on their level of pleasantness, arousal, and dominance. In this model, known as the PAD model, emotions are represented within a three-dimensional space formed by the three mentioned dimensions. For example, happiness is located at (P = .81, A = .51, D = .46) and anger at (P = -.51, A = .59, and D = .25). In this manner, when perceived events are evaluated in terms of the PAD dimensions, they can be mapped into the PAD space in order to trigger a corresponding emotion. The PAD model has been previously used before in [Becker-Asano and Wachsmuth 2009] and [Gebhard 2005]. We leverage the framework designed by [Gebhard 2005] into our proposed AMP model.

3.1.3. Mood

We believe that mood is an important human characteristic that needs to be leveraged when developing a CME. Integrating mood can contribute to the emotion generation mechanism thereby finding applications in situations such as behavioural training of young children [Wood et al. 2021]. However, most of the CMEs for robots do not integrate the notion of mood [Ojha et al. 2020]. We leverage the PAD model and ALMA to integrate the characteristic of mood into our framework. According to [Mehrabian 1996b], by separating each axis into positive and negative, the eight resulting regions in the PAD space can be labeled as follows:

- +P+A+D Exuberant
- -P-A-D Bored
- +P+A-D Dependent
- -P-A+D Disdainful
- +P-A+D Relaxed
- -P+A-D Anxious
- +P-A-D Docile
- -P+A+D Hostile

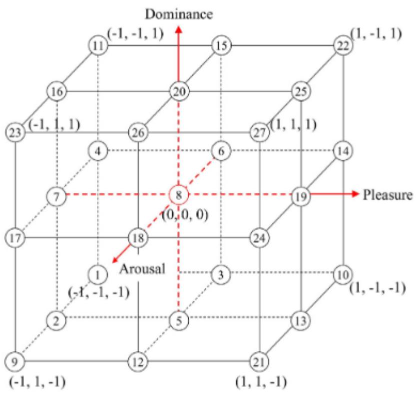

Figure 7. Representing Mehrabian's PAD space [Yan et al. 2021]

The octants are depicted in Figure 7. These eight subspaces serve as mood octants in ALMA [Gebhard 2005] and we adopt the same.

3.1.4. Personality

We also identify personality as another important human characteristic to be implemented into our CME. When social robots are intended to be deployed for interacting with people with various background, belief, culture and nature, these robots should be able to adapt their personality. This is where the need of personality comes into play for tuning the emotional and behavioural mechanisms of the robot. Although some CMEs implement the notion of personality, most of them do not [Ojha et al. 2020]. [Mehrabian 1996b] employs the PAD model to additionally represent temperament scales, which makes it possible to define and describe personality types. In this context, points in the PAD space determine individual traits, regions define personality types, and lines that cross the intersection of the axes define particular dimensions of personality [Rodríguez and Corchado 2013]. From core psychology research, there are many personality models that consist of a set of dimensions, where every dimension is a specific property of the personality. The widely accepted among them are the Big Five Personality Traits [McCrae and John 1992], also called the OCEAN model (see Figure 8) . We combine both these models in a similar manner to [Gebhard 2005] to integrate personality traits in our framework.

Factor	Description	Adjectives used to describe
Openness	Open minded-ness, interest in culture	Imaginative, creative, explorative
Conscientiousness	Organized, persistent in achieving goals	Methodical, well organized, dutiful
Extraversion	Preference for and behaviour in social situations	Talkative, energetic, social
Agreeableness	Interactions with others	Trusting, friendly, cooperative
Neuroticism	Tendency to experience negative thoughts	Insecure, emotionally distressed

Figure 8. OCEAN model of personality [Egges et al. 2004]

3.1.5. Domain Independence

CMEs in social robots should be able to process and exhibit emotional responses in various situations. Therefore, if a model is designed to work only in certain kind of interaction domain, it limits the applicability of the model in real world human - robot interactions. In accordance with the arguments of [Gratch and Marsella 2004], we design our model in such a way that it is able to adopt the flexibility to function in a domain-independent manner. We do this via a combination of data-driven approach and using multi-target regression [Tsoumakas et al. 2014] to generate appraisal dimensions.

3.1.6. Data Driven

According to [Ojha et al. 2020], an emotion model based on appraisal theory should map the appraisals (situation assessments) into corresponding emotion intensities as defined by the theory. Since most appraisal theories do not provide a concrete quantifiable mapping rules from appraisals to emotion, most computational implementations adopt ad-hoc approaches to support their research goals. [Ojha et al. 2020] suggest that such a mapping should be done in a data driven manner using machine learning techniques and we employ the same as mentioned earlier in the previous subsection.

3.2. Key Inspirations for the AMP Framework

3.2.1. Multi-Target Regression as Emotion Appraisal

The inspiration for using multi-target regression for appraisal was drawn from [Ong et al. 2021]. The authors of [Ong et al. 2021] rightly point out that research in affective computing has traditionally fallen into either theory-driven approaches that may not scale well to the complexities of naturalistic data, or atheoretic, data-driven approaches that learn to recognize complex patterns but still fall short of reasoning about emotions. They introduce deep probabilistic programming, a new paradigm that models psychologically-grounded theories of emotion using stochastic programs. Specifically, their framework combines the benefits of probabilistic models with those of deep learning models, combining the advantages of both approaches. They provide an example of how when modelling someone's emotions in a specific context, one may choose to compose a probabilistic model of emotional appraisal with a deep learning model for recognizing emotions from faces and do this all within a single unified framework for training and performing inference. The authors further state that by leveraging modern advances in deep probabilistic programming languages, researchers can easily scale these models up to larger, naturalistic datasets.

The idea of using regression for emotion appraisal was borrowed from one of their tutorials. The tutorial explains how probabilistic programming can be utilised to perform regression as emotional appraisal. While this would certainly be the most optimal way to proceed with, the authors do not show how this idea could be extended to multi-target tasks. This is quite essential as events in a person-environment sphere are generally mapped to not one but multiple appraisal variables and using only one variable would be a huge handicap. However, we retain the idea of using regression for appraisal and build our own feed-forward neural network to perform multi-target regression. Figure 9 represents the process of regression as appraisal. β_i denotes the weights for the regression model. They are a depiction of the model being used in the process. The models could either be based on probabilistic inference or any other approaches such as deep learning.

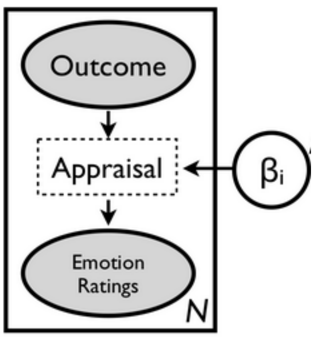

Figure 9. Regression as Appraisal [Ong et al. 2021]

3.2.2. ALMA

ALMA relies on an OCC based model [Ortony et al. 1988] called *EmotionEngine* for emotion appraisal. We instead opt for a regression based approach to make our CME domain independent and data driven. [Gebhard 2005]'s ALMA reiterates Mehrabian's definition of mood as "an average of a person's emotional states across a representative variety of life situations". Mood is distinguished from emotions by the fact that it reflects an individual's stable or longer lasting affective state as compared to short-term effects of emotion. We adopt this implementation of Mood into our framework.

3.3. Integrating and Building the AMP Framework

Although a number of computational model of emotions implement modules representing *mood*, *personality* and *emotions*, they don't necessarily capture the relationship between them. The model proposed by [Egges et al. 2004] is one of the few models to address this and provide an architecture comprehensive enough to capture the mutual interaction among personality factors, mood and emotion. We take inspiration from their model to establish the interconnection between the influence of emotion, mood and personality on each other.

In our model, we adopt a data-driven approach to make the model flexible enough to be domain independent. The appraisal process in our model results in the initial activation of emotions. The activated emotions then influence the current mood to be generated. The mood is also influenced by the personality traits that the model is initialized with. The generated mood in turn influences the final emotional intensities to be generated thus resulting in an *emotion-mood-emotion* cycle. The interconnection between the influence of each component on each other is depicted in Figure 10 .

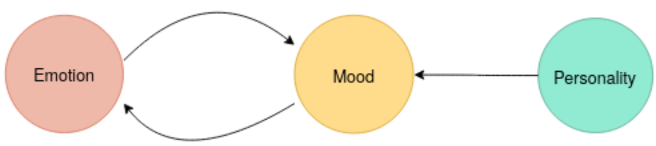

Figure 10. Interconnection between Emotion, Mood and Personality

In this section, we describe in detail the implementation of each component in our CME. The architectural diagram for AMP can be referred in Figure 11 .

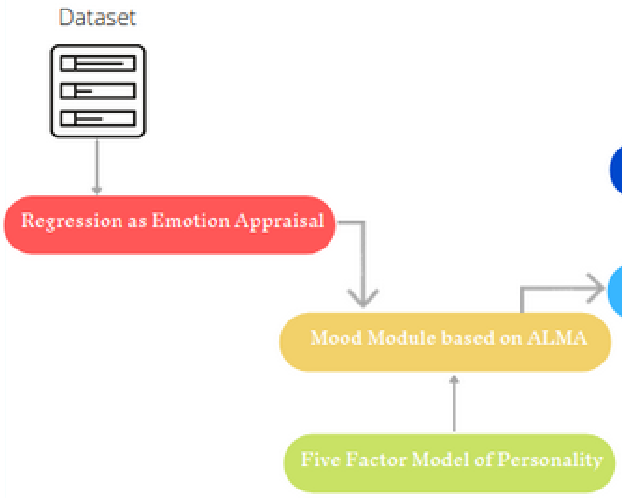

Figure 11. AMP Architecture

3.3.1. Dataset

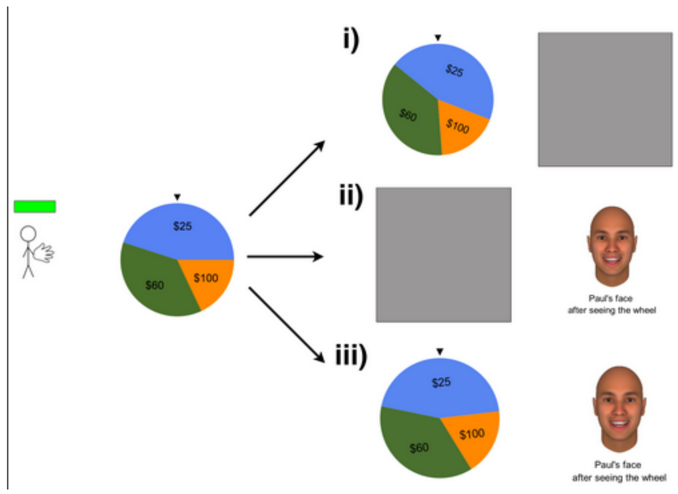

Figure 12. Game played in the experiment [Ong et al. 2015]

The data that we would be using for building and initial experimentation of our model is taken from ([Ong et al. 2015]; Experiment 3), which is available here. In this experiment, [Ong et al. 2015] showed human participants an agent playing a gamble (refer to Figure 12); he spins a wheel with three possible outcomes, and wins the amount on the wheel. The participants were shown three kinds of trials :

- On some trials, participants see the outcome that the agent won ((i) in the figure).
- On other trials, participants were not shown the outcome, but instead were shown what ostensibly was the agent's facial expression after seeing the outcome ((ii) in the figure).
- On the last third of trials, participants were shown both the outcome and the agent's facial expression ((iii) in the figure).

Following these, participants were asked to rate how they thought the agent felt, on 8 emotions, using a 9 point Likert scale [Likert 1932]. Thus, the dataset consists of some "outcome only" trials where participants saw outcomes and rated the agent's emotions, "facial expression only" trials where participants attributed emotions to a facial expression, and trials where they saw both and had to integrate the information from both the outcome and the facial expression to make a judgment.

The data is stored as a torch Tensor of size (1541, 17), indicating that there are N=1,541 observations of 17 variables. The first 9 are the parameterization of the outcome (the 3 payoffs on the wheel and their probabilities, which outcome they won and that probability, and the angle within the sector that the wheel landed on), and the next 8 are the emotion variables. All the variables are scaled so that they lie within [0,1]. For our purposes, we would be working with only 7 of the 8 emotion variables provided, the reason for which is given in a later subsection. A sample image of the dataset is given in Figure 13.

```
Reading in dataset...
Preview of first 3 rows:
   payoff1  payoff2  payoff3  prob1  prob2  prob3  win  winProb  angleProp
0     0.50     0.75      0.9   0.30   0.52   0.18  0.5     0.30      0.921
1     0.15     0.70      0.8   0.45   0.29   0.26  0.8     0.26      0.873
2     0.50     0.75      0.9   0.30   0.52   0.18  0.5     0.30      0.467
   happy    sad  anger  disgust  fear  content  disapp
0  0.625  0.000  0.000     0.25   0.0    0.625   0.375
1  0.875  0.000  0.000     0.00   0.0    0.000   0.000
2  0.625  0.125  0.125     0.00   0.0    0.250   0.500
```

Figure 13. 3 Sample Rows of the Dataset. The above 9 variables represent the input while the lower half represents the corresponding scaled Likert scale outputs.

3.3.2. Emotion Appraisal

As mentioned earlier, we use multi-target regression to perform emotional appraisal. We build a simple fully connected neural network to perform this task (see Figure 13). The network takes in the 9 parameters representing the outcome of a trial of the game as input. The outcome layer comprises of 7 units, each depicting the intensity of the corresponding emotion it represents. In between, the network comprises of four hidden layers of 320, 160, 80 and 40 units each. A sufficiently wide and moderate depth for the network was required due to multiple targets (7) being predicted at a time.

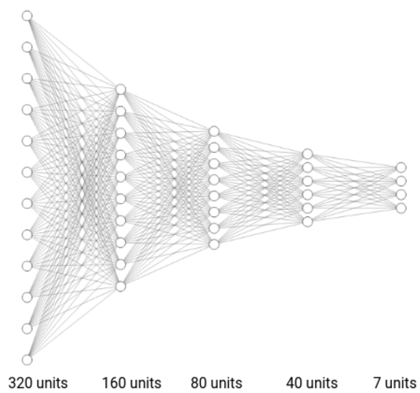

Figure 14. Neural Network Architecture

We use the ReLU activation function in the hidden layers. The model is fit using mean absolute error (MAE) loss and the Adam version of stochastic gradient descent.

We also use k-fold cross-validation to get an unbiased estimate of model performance. This is feasible as there is not too much data and the process can be completed in a reasonable time. The model is evaluated on the multi-output regression task using repeated k-fold cross-validation with ten folds and three repeats. In each fold, the model is defined, fit, and evaluated. The scores are collected and can be summarized by reporting the mean and standard deviation. In our model, the MAE achieved is around 0.131.

3.3.3. Personality-Mood Influence

Before modeling the effects of emotions on mood, it is essential to define an individual's default mood as a mood start value. [Mehrabian 1996a] presents a reliable mapping between the big five personality traits and the PAD space as :

$$
\begin{aligned}
Pleasure &= (0.21 * Extraversion) + (0.59 * Agreeableness) + (0.19 * Neuroticism) \\
Arousal &= (0.15 * Openness) + (0.30 * Agreeableness) - (0.57 * Neuroticism) \\
Dominance &= (0.25 * Openness) + (0.17 * Conscientiousness) \\
&\quad + (0.60 * Extraversion) - (0.32 * Agreeableness)
\end{aligned}
$$

This helps us in setting up an individual's default mood value in the PAD space. Personality, thus, governs the value that a mood would begin with and revert to after extended periods of inaction. The factors of personality are not influenced by anything as they are considered to be fixed after adulthood [McCrae and Jr. May]. where e_i represents the emotion intensities computed from the appraisal process and P, A, D represents that the weighted averages are computed for each coordinate separately. We also note that we use 7 emotion intensities out of the available 8. This is because no mapping or alternate mapping was available for the emotion *Surprise* and we thus decided to forego it.

3.3.4. Emotion-Mood Influence

After the appraisal process yields the initial activated emotions , it is now time to compute the current mood based on it. We use ALMA's [Gebhard 2005] *pull and push mood change function* to model the emotion-mood influence in AMP.

We first begin by representing all emotions available in our dataset in the PAD space. The mapping from [Gebhard 2005] has been utilised to do this (see Figure 15). Direct mappings were not available for all emotions in our dataset and we use synonymous emotions available in the mapping for these cases. These include using the mapping for Joy to represent Happy, Remorse to represent Sad, Shame to represent Disgust and Gratification to represent Content.

We then proceed with the computation of the Virtual Emotion Centre (VEC) of all active emotions in the PAD space. The VEC also represents a point in the PAD space and is computed as a weighted average of all active emotions:

$$VEC_{P,A,D} = \frac{\Sigma_{i=1}^{7}(e_i * Emotion_{P,A,D})}{\Sigma_{i=1}^{7} e_i}$$

where e_i represents the emotion intensities computed from the appraisal process and P, A, D represents that the weighted averages are computed for each coordinate separately. We also note that we use 7 emotion intensities out of the available 8. This is because no mapping or alternate mapping was available for the emotion *Surprise* and we thus decided to forego it.

Emotion	P	A	D	Mood Octant
Admiration	0.5	0.3	-0.2	+P+A-D Dependent
Anger	-0.51	0.59	0.25	-P+A+D Hostile
Disliking	-0.4	0.2	0.1	-P+A+D Hostile
Disappointment	-0.3	0.1	-0.4	-P+A-D Anxious
Distress	-0.4	-0.2	-0.5	-P-A-D Bored
Fear	-0.64	0.60	-0.43	-P+A-D Anxious
FearsConfirmed	-0.5	-0.3	-0.7	-P-A-D Bored
Gloating	0.3	-0.3	-0.1	+P-A-D Docile
Gratification	0.6	0.5	0.4	+P+A+D Exuberant
Gratitude	0.4	0.2	-0.3	+P+A-D Dependent
HappyFor	0.4	0.2	0.2	+P+A+D Exuberant
Hate	-0.6	0.6	0.3	-P+A+D Hostile
Hope	0.2	0.2	-0.1	+P+A-D Dependent
Joy	0.4	0.2	0.1	+P+A+D Exuberant
Liking	0.40	0.16	-0.24	+P+A-D Dependent
Love	0.3	0.1	0.2	+P+A+D Exuberant
Pity	-0.4	-0.2	-0.5	-P-A-D Bored
Pride	0.4	0.3	0.3	+P+A+D Exuberant
Relief	0.2	-0.3	0.4	+P-A+D Relaxed
Remorse	-0.3	0.1	-0.6	-P+A-D Anxious
Reproach	-0.3	-0.1	0.4	-P-A+D Disdainful
Resentment	-0.2	-0.3	-0.2	-P-A-D Bored
Satisfaction	0.3	-0.2	0.4	+P-A+D Relaxed
Shame	-0.3	0.1	-0.6	-P+A-D Anxious

Figure 15. Mapping Emotions to the PAD Space [Gebhard 2005]

We now proceed to implement [Gebhard 2005]'s push and pull mood change function in our context. The working of the function is shown in Figure 16. If the current mood position is between the PAD space's zero point and the VEC, the current mood is attracted towards the VEC. This is called the pull phase. If the current mood is beyond (or at) the VEC the current mood is pushed away, further into the current mood octant in which the mood is located. This is called the push phase. The push phase realizes the concept that a person's mood gets more intense the more experiences he/she faces that support this mood. The intensity by which the current mood is attracted towards or pushed away from VEC is computed as:

$$PullPushIntensity = \frac{\Sigma_{i=1}^{7} e_i}{7.0}$$

where e_i represents the emotion intensities computed during appraisal.

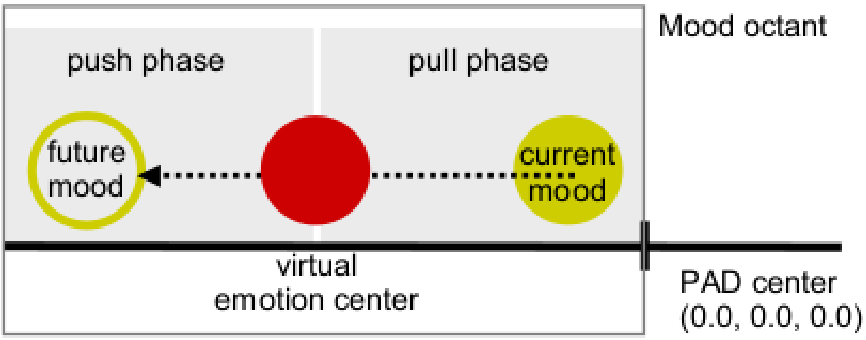

Figure 16. Pull and Push Mood Change Function [Gebhard 2005]

The pseudocode for our implementation of the pull and push function is given as follows:

Algorithm 1 MOOD-EMOTION INFLUENCE ALGORITHM

---

dist = Distance between VEC and CurrentMood in PAD space
scale = *PullPushIntensity* / dist

if *Both VEC and CurrentMood lie in the same octant* **then**
  if $VEC_P >= 0$ **then**
    if $CurrentMood_P < VEC_P$ **then**

$$NewMood_{P,A,D} = CurrentMood_{P,A,D} + (scale * (VEC_{P,A,D} - CurrentMood_{P,A,D}))$$

    else

$$NewMood_{P,A,D} = CurrentMood_{P,A,D} + (scale * (CurrentMood_{P,A,D} - VEC_{P,A,D}))$$

    end
  end
  if $VEC_P < 0$ **then**
    if $CurrentMood_P > VEC_P$ **then**

$$NewMood_{P,A,D} = CurrentMood_{P,A,D} + (scale * (VEC_{P,A,D} - CurrentMood_{P,A,D}))$$

    else

$$NewMood_{P,A,D} = CurrentMood_{P,A,D} + (scale * (CurrentMood_{P,A,D} - VEC_{P,A,D}))$$

    end
  end
end

Note that the pseudocode only gives the implementation for when both VEC and CurrentMood lie in the same octant. For all other cases, the new mood is computed simply as :

$$NewMood_{P,A,D} = CurrentMood_{P,A,D} + (scale * (VEC_{P,A,D} - CurrentMood_{P,A,D}))$$

Also note that the use of P, A, D is to indicate that each computation is done individually across each dimension.

3.3.5. Mood-Emotion Influence

Computation of the new mood value allows us to generate the final intensities for the emotions activated. [Gockley et al. 2006] have noted the use of the following equation to model the influence of mood on emotions and we use the same:

$$NewEmotion_{P,A,D} = OriginalEmotion_{P,A,D} * (1 + (0.25 * Val_{P,A,D} * NewMood_{P,A,D}))$$

where, $Val_{P,A,D}$ is 1 if $sgn(OriginalEmotion_{P,A,D}) = sgn(NewMood_{P,A,D})$ else -1

The intensity of the generated emotion is computed by calculating the norm of the emotion vector in PAD space. However, one must note that the norm calculated cannot be used as is to represent the emotion intensity. This is because, when mapping the original emotion intensities onto the PAD space, we did not take into account the contribution to intensity generated by the original representation in the PAD space, i.e, we need to scale the new norm. This is done as follows :

$$FinalEmotionIntensity_i = \frac{||NewEmotion_i(P, A, D)|| * e_i}{||OriginalEmotion_i(P, A, D)||}$$

where e_i are the emotion intensities generated from the appraisal process.

The codebase to run the proposed computational model of emotion can be found here.

4. Conclusion

We were thus able to develop a compact but powerful computational model of emotion. We used a game based experimental dataset for the development and testing of our model. This game-based methodology was particularly chosen to show how our model can be deployed flexibly across domains. Serious games are an important and popular approach to facilitate improved learning experiences among autistic children [Zakari 2014]. If a social robot is to engage in a serious game with a child on the autism spectrum disorder, it needs to have the capability to generate emotions for empathetic interactions. Any game that can be parameterized based on its inputs, rewards and features would fit very well with our proposed model.

Since the current model has not been tested against real-time interactions, it lacks time-dependent emotion and mood decay functions which would be essential for looped interactions. However, these can be easily incorporated into the model as all components are modular and adding new components to build upon the model would be fairly easy.

References

[Becker-Asano and Wachsmuth 2009] Becker-Asano, C. and Wachsmuth, I. (2009). Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems*, 20:32–49.

[Breazeal 2009] Breazeal, C. (2009). Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364:3527 – 3538.

[Chen and Poo 2003] Chen, P. C. and Poo, A.-N. (2003). Engineering, artificial intelligence in. In Bidgoli, H., editor, *Encyclopedia of Information Systems*, pages 141–155. Elsevier, New York.

[Egges et al. 2004] Egges, A., Kshirsagar, S., and Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents. *Journal of Visualization and Computer Animation*, 15:1–13.

[Frijda et al. 1989] Frijda, N., Kuipers, P., and ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness.

[Gebhard 2005] Gebhard, P. (2005). Alma: a layered model of affect. pages 29–36.

[Gockley et al. 2006] Gockley, R., Simmons, R., and Forlizzi, J. (2006). Modeling affect in socially interactive robots. In *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 558–563.

[Gratch and Marsella 2004] Gratch, J. and Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cogn. Syst. Res.*, 5(4):269–306.

[Höök 2009] Höök, K. (2009). Affective loop experiences: designing for interactional embodiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364:3585 – 3595.

[Hung et al. 2019] Hung, L., Liu, C., Woldum, E., Au-Yeung, A., Berndt, A., Wallsworth, C., Horne, N., Gregorio, M., Mann, J., and Chaudhury, H. (2019). The benefits of and barriers to using a social robot paro in care settings: A scoping review. *BMC Geriatrics*, 19.

[Likert 1932] Likert, R. (1932). A technique for the measurement of attitudes. In *Archives of Psychology*.

[Marsella et al. 2010] Marsella, S., Gratch, J., and Petta, P. (2010). Computational models of emotion. *A Blueprint for Affective Computing-A Sourcebook and Manual*, pages 21–46.

[McCrae and John 1992] McCrae, R. and John, O. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, 60 2:175–215.

[McCrae and Jr. May] McCrae, R. and Jr., P. C. (1988 May). Personality in adulthood: a six-year longitudinal study of self-reports and spouse ratings on the neo personality inventory. *J Pers Soc Psychol*, 54(5):853–63.

[Mehrabian 1996a] Mehrabian, A. (1996a). Analysis of the big-five personality factors in terms of the pad temperament model. *Australian Journal of Psychology*, 48(2):86–92.

[Mehrabian 1996b] Mehrabian, A. (1996b). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current psychology (New Brunswick, N.J.)*, 14:261–292.

[Mitsunaga et al. 2008] Mitsunaga, N., Smith, C., Kanda, T., Ishiguro, H., and Hagita, N. (2008). Adapting robot behavior for human–robot interaction. *Robotics, IEEE Transactions on*, 24:911 – 916.

[Myers 2004] Myers, D. (2004). Theories of emotion. *Psychology*, 7.

[Ojha et al. 2020] Ojha, S., Vitale, J., and Williams, M.-A. (2020). Computational emotion models: A thematic review. *International Journal of Social Robotics*.

[Ong et al. 2021] Ong, D. C., Soh, H., Zaki, J., and Goodman, N. D. (2021). Applying probabilistic programming to affective computing. *IEEE Transactions on Affective Computing*, 12(2):306–317.

[Ong et al. 2015] Ong, D. C., Zaki, J., and Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition*, 143:141–162.

[Ortony et al. 1988] Ortony, A., Clore, G., and Collins, A. (1988). *The Cognitive Structure of Emotion*, volume 18.

[Ortony and Turner 1990] Ortony, A. and Turner, T. (1990). What's basic about basic emotions? *Psychological review*, 97:315–31.

[Paiva et al. 2014] Paiva, A., Leite, I., and Ribeiro, T. (2014). Emotion modelling for social robots.

[Pino et al. 2015] Pino, M., Boulay, M., Jouen, F., and Rigaud, A. (2015). "are we ready for robots that care for us?" attitudes and opinions of older adults toward socially assistive robots. *Frontiers in Aging Neuroscience*, 7:141.

[Rodríguez and Corchado 2013] Rodríguez, L.-F. and Corchado, F. (2013). Development of computational models of emotions for autonomous agents: A review. *Cognitive Computation*, 6:351–375.

[Roseman et al. 1990] Roseman, I., Spindel, M., and Jose, P. (1990). Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. *Journal of Personality and Social Psychology*, 59:899–915.

[Russell and Mehrabian 1977] Russell, J. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11:273–294.

[Scherer 2001] Scherer, K. (2001). Appraisal considered as a process of multilevel sequential checking.

[Spezialetti et al. 2020] Spezialetti, M., Placidi, G., and Rossi, S. (2020). Emotion recognition for human-robot interaction: Recent advances and future perspectives. *Frontiers in Robotics and AI*, 7:145.

[Tsoumakas et al. 2014] Tsoumakas, G., Spyromitros-Xioufis, E., Vrekou, A., and Vlahavas, I. (2014). Multi-target regression via random linear target combinations. In Calders, T., Esposito, F., Hüllermeier, E., and Meo, R., editors, *Machine Learning and Knowledge Discovery in Databases*, pages 225–240, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Wood et al. 2021] Wood, L. J., Zaraki, A., Robins, B., and Dautenhahn, K. (2021). Developing kaspar: A humanoid robot for children with autism. *International Journal of Social Robotics*, pages 1–18.

[Yan et al. 2021] Yan, F., Iliyasu, A. M., and Hirota, K. (2021). Emotion space modelling for social robots. *Engineering Applications of Artificial Intelligence*, 100:104178.

[Zakari 2014] Zakari, H. (2014). A review of serious games for children with autism spectrum disorders (asd).