An Affective Model for the Construction of a Computer Chat Agent

Charles S. Henry

Abstract

Emotions play a critical role in human cognition and are an integral part of the human experience. They serve an important role in decision-making, perception, human interaction, and intelligence. It can be argued that in order to be intelligent one must exhibit emotions. This work explores one aspect of ‘affective computing’, computing that relates to, arises from, or deliberately influences emotions. A model for an emotionally intelligent computer chat agent is outlined in this paper. This model will use the five-factor model of personality derived from psychological studies to construct the agent’s personality. This agent will communicate using natural language and will also recognize the emotions of the user and respond according to its personality. It will be shown that the cognitive agent represents a more human-like chat partner.

1. Introduction

The movie “2001: A Space Odyssey” featured a computer named HAL that interacted with the crew of the spaceship Discovery. This computer carried on conversations, played chess, and even conducted an interview with a news agency. The creation of an artificially intelligent computer has been envisioned since the creation of computing machines.

Being intelligent is difficult to define. Is being intelligent a single faculty or is it a collection of distinct and unrelated abilities? Does one have to be self-aware to be intelligent? Are computers capable of exhibiting intelligence? Lady Lovelace’s objection as noted in [5] states computers can only do as they are told and consequently cannot perform original (hence, intelligent) actions. However, computers can give the illusion of being intelligent. The Chinese room paradox [9] is a good illustration of this. This paradox has two players, a person inside a room who does not speak Chinese, and a person outside the room who does. The person outside the room asks the person inside the room questions by passing a slip of paper under the door. The person inside the room has a manual that says ‘if you see a squiggle like this, then draw a stroke like that’. The person inside the room then passes the note to the Chinese speaker outside the room. This way the person outside has no idea that the person inside the room does not speak Chinese. The paradox points out that there is no real difference between appearing to understand and really understanding Chinese. This is the nature of artificial intelligence.

In [7] R. Picard et. al. state that scientists have amassed evidence that emotional skills are a basic component of intelligence, especially for learning preferences and adapting to what is important. There is evidence that machines will require at least some emotion to appear intelligent while interacting with people. Reeves and Hass at Stanford theorize that human-computer interaction is inherently natural and social, following the basics of human-human interaction [2]. R. Picard defines emotional intelligence as the ability to “recognize, express, and have emotions, coupled with the ability to regulate these emotions, harness them for constructive purposes, and skilfully handle the emotions of others” [7]. It follows that building an artificially intelligent computer system should consider emotional factors. This brings us to the study of affective computing, or “computing that relates to, arises from, or deliberately influences emotions” [6].

This study incorporates two aspects of affective computing: 1) A computer’s recognition of the emotional state of its user, and 2) the interaction of a user with a computer designed to exhibit a virtual personality. In [7] R. Picard, et. al. developed a system that is able to classify the emotional state of a subject based on physiological data gathered. The results of their study provided 81% recognition accuracy on eight categories of emotion. This study is limited in its scope, it only studied one subject, but the results prove the feasibility of emotional recognition. In [4] S. Kshirsagar, et. al. provide the framework for constructing a chat application that models personality, moods and emotions. This study extends the work of S. Kshirsagar, et. al. to include the recognition of the user’s emotional state.

2. Background

Alan Turing proposed a test, known as the ‘Turing Test’ [8] to measure whether or not a computer is intelligent. In this test a moderator asked a computer and a human questions anonymously through a Teletype. If the moderator was unable to distinguish the difference between a computer and the human then it was concluded that the computer was ‘intelligent’. One of the hurdles in human computer interaction was the lack of emotion, or emotional intelligence. Emotional intelligence consists of the ability to “recognize, express, and have emotions, coupled with the ability to regulate these emotions, harness them for constructive purposes, and skilfully handle the emotions of others” [7]. R. W. Picard linked the importance between computing and emotions in [6]. She believes this a rich field of research important to advancing emotion and cognition theory as well as aiding in the interaction between humans and computers. In order to construct an emotionally intelligent computer system, a basic understanding of the make-up of emotions is required. The following discussion highlights the important components of an emotional personality.

2.1 Personification

Developing a computer application that emulates the human experience requires one to define what it is to appear human. According to the Webster dictionary to represent an inanimate object as having human qualities is to personify. Kshirsagar in [4] worked to develop a virtual human and
chose to break down its personification using four distinct elements.

![Personification Diagram]

**Figure 1. Elements of Personification**

logical personification is the ‘engine’ of the virtual human. This element analyzes input, thinks and generates natural language responses. It can be thought of as the virtual ‘brain’. The emotional component would represent the ‘mind’. It works with the logical component to develop the emotional state and generate appropriate responses. It can be thought that the logical component consults the emotional component on what to do next.

The Physical and Expressional components of personification deal with the actual look of a virtual human. The physical personification would be the facial and body features, for example, the personification could be male or female, large or small stature. The expressional personification deals specifically with the physical animations that result from an emotional change or state. For example, if the virtual human were happy one would expect to see a smile on her face. These personifications can be modeled using 3D modeling software and could consist of a library of displays and transformations. This study does not deal with the physical or emotional personification, instead concentrating on the underlying emotional and logical aspects of the virtual personality.

### 2.2 Personality

Psychological research has proposed a Five Factor Model (FFM) of personality to characterize an individual’s personality [4]. Not only is this model used for a general understanding of personality, psychologists also utilize it when identifying personality disorders. These factors are important in personifying a virtual human, they directly influence the type of responses generated by the logical/emotional components of the application.

### 2.3 Emotions

Emotions are defined as “a mental state that arises spontaneously rather than through conscious effort and is often accompanied by physiological changes; a feeling.”[10] They can be characterized as reactions to events, actions and objects. Many emotional reactions are physiological, such as a smile on an individual’s face, or a rapid heartbeat when startled.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Description</th>
<th>Imagery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boredom</td>
<td>No Emotion, Vacancy</td>
<td>Blank paper, Typewriter</td>
</tr>
<tr>
<td>Anger</td>
<td>Desire to fight</td>
<td>People who arouse rage</td>
</tr>
<tr>
<td>Hate</td>
<td>Passive anger</td>
<td>Injustice, Cruelty</td>
</tr>
<tr>
<td>Grief</td>
<td>Loss, Sadness</td>
<td>Deformed child, Loss of mother</td>
</tr>
<tr>
<td>Platonic Love</td>
<td>Happiness, Peace</td>
<td>Family, Summer</td>
</tr>
<tr>
<td>Romantic Love</td>
<td>Excitement, Lust</td>
<td>Romantic encounters</td>
</tr>
<tr>
<td>Joy</td>
<td>Uplifting happiness</td>
<td>Music to &quot;Ode to Joy&quot;</td>
</tr>
<tr>
<td>Reverence</td>
<td>Calm, Peace</td>
<td>Church, Prayer</td>
</tr>
</tbody>
</table>

There is no common agreement as to the definition of basic emotions. Ortony, Clore and Collins proposed a model (known as the OCC model) that categorizes 22 positive or negative reactions [4]. However, not all these emotions are known to elicit a physical response. “Facial recognition is easier for people, e.g., 70-98% accurate on six categories of facial expressions exhibited by actors [7].” This limits the number of emotions an affective computer system will be able to perceive. In order to accommodate emotion recognition six emotions from the OCC model that elicit a physical response were chosen for this exercise, they are listed along with their corresponding mood in table 3.
Table 3. Emotions used as emotional tags and corresponding Mood

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Adjectives</th>
<th>Mood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy</td>
<td>Happy-for, Gloating, Pride, Admiration, Love, Hope, Relief, Gratification, Gratitude</td>
<td>Good</td>
</tr>
<tr>
<td>Sadness</td>
<td>Resentment, Pity, Distress, Shame, Remorse</td>
<td>Good</td>
</tr>
<tr>
<td>Anger</td>
<td>Reproach, Hate</td>
<td>Bad</td>
</tr>
<tr>
<td>Surprise</td>
<td>Startled, Excited</td>
<td>Neutral</td>
</tr>
<tr>
<td>Fear</td>
<td>Scared, Anxious</td>
<td>Bad</td>
</tr>
<tr>
<td>Disgust</td>
<td>Disappointed, Offended</td>
<td>Neutral</td>
</tr>
<tr>
<td>Neutral</td>
<td>Bored, Vacant</td>
<td>Neutral</td>
</tr>
</tbody>
</table>

3. System Overview

The goal of this system is to emulate the dialogue that a person would have in an Internet chat room with the addition of the participants being able to recognize each others emotional response, similar to what would happen if there were a video feed between the participants. The interaction consists of a series of questions, requests, and responses. Since currently there is no way to determine a persons emotional state in real time via a video feed the users emotional state will be an input in addition to the chat dialogue. The system used to implement an affective chat agent is described in the following sections. It essentially consists of three major modules that take the input chat string (with the users emotion input), process it to determine the potential responses, consult the personality processing module, determine the emotion / mood change, and then outputs an appropriate response (including emotional tag). A detailed description of each module follows.

2.4 Mood

In order to link emotional expression and the personality the concept of mood is introduced. There is a clear separation between mood and personality. “Personality causes deliberative reactions, which in turn causes the mood to change [4].” Moods and emotions are only differentiated in terms of levels of arousal. Emotions are much more spontaneous and elicit a physiological response where the mood is an underlying state of being. Moods influence the emotional response of the chat-robot, on the other hand, and the nature of the interaction with the human participant influence the mood of the chat-robot. This study limits the possible moods to good, neutral or bad.

The relationship between Personality, Mood and Emotion can be summarized as follows [4]:
1. Personality practically does not change over time. It causes deliberative reaction and effects how moods change in a dialogue over time.
2. Mood, from a higher level is affected by the personality, and it is also affected from the lower level by the emotional state.
3. On the lowest level, the instantaneous emotional state is influenced by mood as well as the current dialogue state.

Figure 2. Relationship of Personality, Mood and Emotion.

Figure 3. System Overview
3.1 Text Processing and Response Generation

This module takes the input text and generates an appropriate response. It is based on ALICE (Artificial Linguistic Internet Computer Entity) [1], an open source project developed by Dr. Richard S. Wallace. ALICE uses AIML (Artificial Intelligence Markup Language) to interface with a database of possible responses to user inputs. The AIML standard consists of a basic unit called a category. Each category has an input question, an output answer, and an optional context. The question uses a <pattern> tag, and answers use a <template> tag. Context is captured using the tags <that> and <topic>. The <that> tag appears inside the <category> and its pattern must match the chat-robot’s last utterance, while the <topic> tag appears outside the <category> tag and is used to group categories together. If there is no <that> tag it is assumed to be a wildcard. By matching the <that> tag the dialogue is sure to be coherent and within context. The following simple example of AIML is activated when the client types KNOCK KNOCK:

```
<category>
  <pattern> KNOCK KNOCK </pattern>
  <template> Who is there? </template>
</category>
```

The chat-robot would respond with “Who is there?” An example of a random response in AIML follows:

```
<category>
  <pattern> HOW ARE YOU </pattern>
  <template>
    <random>
      <li> I am doing very well. How are you? </li>
      <li> Not so well today. </li>
      <li> Everything is running smoothly. </li>
    </random>
  </template>
</category>
```

In this example the chat-robot would randomly choose one of the responses with the <li> tag.

AIML does not employ any syntactic or semantic language analysis techniques. However, the features included in AIML make it more than just a pattern matching application. Being able to respond randomly and within context makes it a believable chat participant. It is limited to its knowledge base of possible responses. Analysis of the human input to ALICE found that the set of things people could possibly say which are grammatically correct or semantically meaningful is surprisingly small. Data indicates that 1800 words cover 95% of all the first word inputs to ALICE, and the average number of second word choices is two. According to Wallace in [9] 6,000 patterns cover 95% of all recorded inputs to ALICE.

The intent of this application is to express and recognize emotions within dialogue. The authors of [4] proposed to extend AIML to include emotional tags in the responses. Each AIML response can be associated with one or more emotion. The chat-robot’s response can then properly be computed and the appropriate emotional and mood change can be processed by the mood to emotion processing module. Following is a simple single response AIML category [4]:

```
<category>
  <pattern> What are you doing </pattern>
  <template>
    <template>
      <emo name="pride" prob="30">
      <emo name="distress" prob="70"> I am very busy now a days. </template>
    </template>
</category>
```

The response in this entry reflects a 30% probability of an emotional response of pride and a 70% probability of distress. This concept can be further implemented to include a variety of responses. In this example the response would have a direct influence on the chat-robot’s mood. To extend on the random response example previously given we can do the following:

```
<category>
  <pattern> HOW ARE YOU </pattern>
  <template>
    <emo name="joy" prob="30" res="I am doing very well. How are you?"/>
    <emo name="sadness" prob="30" res="Not so well today. "/>
    <emo name="neutral" prob="40" res="Everything is running smoothly."/>
  </template>
</category>
```

In this example the application can respond with an emotion corresponding to joy, sadness, or neutral. The personality model takes the input emotion from the user and each of the possible response emotions and consults the mood to emotion processor to choose the most appropriate response.

3.2 Personality Model

The personality model is responsible for changing the mood of the chat-robot. Several factors contribute to a mood change. These include the current mood, input emotion, response mood, and the history of the previous mood decision. When chat is initiated the current mood can be set to good, bad, or neutral, but in most cases it defaults to neutral. The input mood is the human participants emotion converted to mood based on table 3, and the response mood is the possible response emotion again converted to mood based on table 3, if there is no emotion encoded in the response, then it defaults to neutral.

```
Current Mood
\[\text{Input Mood} \rightarrow \text{Response Mood} \rightarrow \text{Changed Mood} \rightarrow \text{Estimated Mood} \rightarrow \text{History} \rightarrow \text{Next Mood} \]

Figure 4. Personality Model
```

The Five Factor Model of the chat-robots personality influences the threshold for mood change. A personality that is mostly agreeable would have a tendency to transition to a
good mood, where a personality that is mostly neurotic would tend to transition to a negative mood. For example, imagine a person in a bad mood but has a personality that is a combination of extroversion and agreeableness. This personality type would tend to get into a good mood quickly. If a conversation begins with a greeting and pleasant topics, this personality would quickly transition to a good mood. The personality model consists of probabilities for mood change, for example a neurotic personality would have a higher probability to change to a bad mood.

The personality model needs to determine the net probability of a mood change based on response probabilities and the mood change probability from the designed personality. This is accomplished through a Bayesian calculation for each of the possible response moods. The following pseudo-code illustrates the calculation of the mood change probability \[4\].

\[
\text{For each emotion } e_i \text{ in response } \]
\[
\begin{align*}
\text{m}_n &= \text{mood corresponding to } e_i \\
P(m_n) &= P(m_n \mid m_i, m_j) \cdot P(e_i)
\end{align*}
\]

The term \(m_n\) is the next mood, \(m_i\) is the current mood and \(P(m_n \mid m_i, m_j)\) denotes the conditional probability for mood change as defined in the personality model. There is a probability associated for each mood, good, bad, and neutral from the personality model, \(P(e_i)\) is the probability associated for each candidate response. If there are multiple possible responses that result in a bad mood, their probabilities are added. The input mood will contribute to the probability for a mood change by simple addition based on a factor defined in the model, this can only increase a probability, not decrease it. For example, if the input mood were negative, only the probability of a negative mood change will be increased. The scale of this contribution is another of the factors that defined the chat-robots virtual personality. A maximum nine probabilities result from this calculation, three possible mood changes times three possible mood responses. A threshold for mood change is also integrated into the model. If \(P(m_n)\) is greater than this threshold, then \(m_n\) would be changed. If this threshold has not been reached, the highest probable mood is stored in history. The application then consults the mood history to see if there is a tendency towards a mood change. The history influences the decision about a probable mood change. For example, assume that the previous response suggested a mood change, but it was below the threshold. The current response also suggests a mood change to the same mood as the previous, but it too did not reach the threshold. The combination of the past mood suggestion and the current mood suggestion would result in a mood change. The determination of which emotional response has yet to be made, that is the responsibility of the mood to emotion processor. The following section discusses the emotional response selection.

#### 3.3 Mood to Emotion Processing

Three factors contribute to the determination of the emotional state: the possible AIML response, the current mood, and the previous emotional state. For each mood, good, neutral and bad, a transition probability matrix is constructed. Again, the values in these tables contribute to the personality of the chat-robot. There is no science as to how these table are populated, only intuition.

Table 4. Emotion transition matrix for good mood.

<table>
<thead>
<tr>
<th>Previous Emotion</th>
<th>Joy</th>
<th>Sadness</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy</td>
<td>0.7</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.05</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.9</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

The first column represents the previous emotion and the first row represents the next expression. For example, if the mood is Good and the current emotion is Neutral, if one of the AIML responses is Joy the transition probability for Joy is 0.1. The following pseudo-code represents how the next mood is chosen \[4\]:

\[
\text{For each emotion } e_i \text{ in response } P(e_i) = \bullet m_n(Ex(e_p),Ex(e_r)) \cdot P(e_i)
\]

\(P(e_i)\) is the probability of an emotion change for each of the possible emotional responses \(e_i\), \(e_p\) is the previous mood. \(m_n(Ex(e_p),Ex(e_r))\) is the lookup factor in the transition probability matrix and \(P(e_i)\) is the probability of the response defined in the AIML tag. The highest \(P(e_i)\) determines the next response, however it does not necessarily mean the emotion has changed. Similar to the mood processor, there is a threshold value that must be met in order to warrant a change in emotion.

#### 3.4 Personality Design

Following is a summary of the components required for designing a virtual personality of a chat-robot. Currently there is no scientific basis for the determination of this personality; human experience and intuition are the only contributing factors.

1. Initial value of the current mood, normally set to neutral.
2. Conditional probabilities for the personality model, used in the calculation for the mood change probability.
3. Input mood contribution factor used in combination with the mood change probability.
4. Threshold value for a mood change.
5. Transition probability matrices for the emotional change, one for good, neutral and bad mood.

#### 4. Results and Conclusion

The sheer scope of encoding the human lexicon with emotional tags is excruciating and not particularly feasible for the scope of this study. Even the library of categories available with ALICE is daunting. As such the concentration of this work was on implementation of the system, and demonstrate its capability to affectively acknowledge emotion and change moods. Unfortunately due to the limited number of categories implemented with emotional tags, the possible responses were predictable and not diverse enough to appear emotionally intelligent. However, there is opportunity to automate the inclusion of emotional tags. ALICE has the ability to be ‘trained’. This is called targeting. Following is an example:
This method of targeting could be extended to include the input of emotional tags. Another issue in the design of an affective chat-robot is that there is no clear definition of what consists of a personality.

This system is the framework for implementation in a number of applications. It could be used as a virtual teacher that could react to emotions being exhibited by a student. It could also be used as a virtual office assistant, and certainly a source of entertainment. One of the potential benefits of this work would be in the study of cognition and psychology. Researchers could use this system to design personalities and study the interaction with human chat partners. Further extension is possible by incorporating a speech and sight recognition system and it could be the ‘brain’ of a robot.

References


