Computational Models of Affect and Empathy for Pedagogical Virtual Agents

Jennifer Sabourin, Bradford Mott, and James Lester

Department of Computer Science, North Carolina State University, Raleigh NC 27695
{jilrobiso, bwmott, lester}@ncsu.edu

Abstract. Evidence of the importance of affective support for students has fueled recent work in designing empathetic virtual agents for virtual learning environments. However, because there is no uniformly accepted model of “learning” emotions, or of empathy, it is difficult to design models of affective agent behavior that are soundly based in psychological research. This paper presents work on several components of an affective reasoning system designed for empathetic agents in the virtual learning environment, CRYSTAL ISLAND. This model makes use of several psychologically-grounded theories of affect and empathy. This paper will present the overarching model utilized to explore affective reasoning in CRYSTAL ISLAND as well as validation of the individual components of the system. The results of this work suggest that in lieu of a universal model of human affect and empathy, selection of domain-specific or application-specific psychological models is useful for informing computational models of affective behavior.

Keywords: Empathetic virtual agents, pedagogical agents, virtual learning environments

1 Introduction

Recent years have seen significant advances in cognitive and behavioral models for virtual agents [1,2,3]. Affective modeling in virtual agents is the subject of increasing attention because of its role in motivating users, supporting them through stressful tasks, and increasing users’ abilities to recognize and regulate emotions. Agents with affective capabilities can form social relations with users to motivate them [3], reduce stress levels [4], and teach children to deal with frustration [5] or bullying [2] by eliciting emotional responses. A critical component of each of these social interactions is the ability to use empathy.

Empathy is the expression of emotion based on another’s situation and not merely one’s own [6]. Its expression can demonstrate that the target’s (the recipient of empathetic expression) feelings are understood or shared. For virtual agents to have empathetic responses to human users they must first be able to assess the user’s emotional state and then produce an appropriate response. Both of these tasks require significant affective reasoning capabilities on the part of the virtual agent and can be benefited by consideration of sound psychological models of affect and empathy.
This paper presents work on several components of an affective reasoning system designed for empathetic agents in the virtual learning environment, CRYSTAL ISLAND. This model makes use of several psychologically-grounded theories of affect and empathy. The overarching model utilized to explore affective reasoning in CRYSTAL ISLAND as well as validation of the individual components of the system is the focus of the following discussion.

2 Background

Empathetic approaches to responding to user affect have been shown to be effective at altering the affective state of the user as well as increasing feelings of interest, motivation, and liking of the virtual environment [5,7,8]. Recent work with empathetic synthetic agents has explored their affective responsiveness to biofeedback information in high stress scenarios [4]. Other work examines how agent’s interpersonal behaviors can be used to encourage social bonding with the user [2,9]. While empathetic virtual agents can help to improve user affect in a variety of settings, a large focus has been developing empathetic agents for computer-based learning environments.

2.1 Empathy in Computer-Based Learning Environments

Learning is a complex and cognitively demanding task which can produce a variety of affective responses in students [10,11,12]. The ability to scaffold the emotional experience of students in the support of overall learning gains is considered a cornerstone of effective teaching [11] and is consequently an important goal for virtual agents in computerized learning environments. This process involves first assessing the emotional state of the student then providing an appropriate empathetic response [6]. These two lines of work have been actively pursued in computer-based learning environments.

2.1.1 Computational Models of Learner Emotions

To date, many models of affect detection have been developed for use in computer-based learning environments. For instance, a model developed by D’Mello et al. considers facial expressions in terms of action units as well as students’ posture and dialog acts to predict students’ emotions as assessed by expert judges [13]. Similarly, Arroyo et al. have found benefit to multiple channels of physical evidence of affect [14]. By adding features such as facial expressions, skin conductivity, posture, and pressure they were able to account for much more variance over using contextual features of the tutoring environment alone. While these models focus on affect prediction unguided by psychological theories of emotion, Conati and Maclaren have incorporated physical sensors into a complex model based on the OCC theory of appraisal [15]. Though they focus only on a subset of the emotions proposed by OCC
they have used a dynamic Bayesian network to capture many of the complex phenomena associated with appraisal theories. This model estimates student goals based on personal traits and behaviors in the environment as well as evidence from physical feedback channels that further support the model’s prediction. The work presented here builds upon the success of using theoretical models of affect to guide the development of computational models and extends this work by consideration theoretical foundations specific to the learning context.

2.1.2 Empathetic Pedagogical Agents

Recent work on empathetic virtual agents has explored agents who are able to mimic the emotional state of students while expressing motivational statements which provide feedback regarding student’s success and efforts [7]. It has also yielded agents capable of interacting with one another and with the user in a virtual learning environment to elicit empathetic behaviors from its users [2]. A common finding of the use of empathetic agents in virtual learning environments is the positive impact that inclusion of these agents has on student motivation and interest in the learning task [5,7,16].

While many models of empathy are built using direct student self-reports (e.g., [16]) or physiological data (e.g., [4]), the model designed by Paiva et al. [2] is based on psychological theories of affect and empathy. The agents in this virtual environment have general affective responses based on the OCC theory of appraisal. They first deduce the affective states of others and then display affective responses based on several factors (e.g., how much they like the other agent or how similar they feel they are). As with the agent’s general affective responses, the design of this empathetic component of behaviour is informed by psychological research.

2.2 Selected Psychological Models of Affect and Empathy

The approaches to affective support used in CRYSTAL ISLAND are similarly based on psychological models including the widely accepted appraisal theory of human emotions [17], Davis’ model of empathy [6], and a model of learning-focused emotions proposed by Elliot and Pekrun [12].

2.2.1 Appraisal Theory

According to the appraisal theory model of emotions, individuals compare events in the environment to their goals and beliefs to develop an understanding of how these events impact their personal situation [17]. This appraisal results in an emotional state as well as associated action tendencies and physiological responses. Upon experiencing this emotion, individuals are then likely to engage in emotion regulation behaviors (coping). Since the emotional state was determined by an interaction of the environment and the individuals’ beliefs, one of these must be altered in order to attenuate the affective state. This distinction leads to two separate types of coping
strategies: emotion-focused and problem-focused. These strategies attempt to alleviate unpleasant emotional experiences by altering either one’s own beliefs or the external environment, respectively.

2.2.2 Model of Learner Emotions

While there are many frameworks for examining learner emotions, the framework described here builds off appraisal theory by focusing on the types of goals a student has while engaged in learning activities, or goal orientation [18]. Students may either view learning in relation to performance or mastery. A performance approach would result in a student wishing to prove his competence and achieve better results than other students. A student with a mastery approach, however, views learning as an attempt to gain a skill, regardless of how her ability compares to others. This distinction between learning and performance goals forms the basis for the appraisal-based theory of learning emotions described by Elliot and Pekrun [12]. This model considers emotions in terms of learning and performance goals, along with evaluations of success and failure in these two categories. Additionally, they argue that certain individuals are more likely to focus on negative or positive valences of achievement emotions. For example, individuals with a positive (approach) disposition are more likely to experience positive feelings of enjoyment and pride, while those with negative (avoidance) dispositions are more likely to experience feelings of anxiety or shame [12]. Though this model does not explicitly reference appraisal theory, its reliance on goals and their evaluation situates it within this framework.

2.2.3 Models of Empathy

The process of empathy involves two distinct phases. First, the emotional situation of the target must be assessed by considering contextual information of the situation or the emotional reactions of the individual [6]. One common feature of many models of empathy is the distinction between judging the emotional state of another by examining the facial expressions and other physical demonstrations of the target or by considering the situation and “simulating” an appraisal of this situation to arrive at an expected emotional state [6,19]. The second step in the empathetic process is the display of empathetic behavior. Here, another distinction is made in how empathetic responses may be given. The empathizer may respond either affectively, by experiencing and expressing an emotional state based on their assessment of the other, or cognitively, resulting in a changed perspective and tendencies towards helping behavior. These two strategies parallel the problem-focused and emotion-focused coping strategies proposed by appraisal theory.
3 Affective Reasoning in CRYSTAL ISLAND

CRYSTAL ISLAND (Figure 1) is a narrative-centered learning environment that is being created in the domain of microbiology for middle school students. It features a science mystery set on a recently discovered volcanic island where a research station has been established to study the unique flora and fauna.

The user plays the protagonist, Alex, who is attempting to discover the source of an unidentified infectious disease at the research station. The story opens by introducing the student to the island and the members of the research team for which her father serves as the lead scientist. As members of the research team fall ill, it is her task to discover the cause and the specific source of the outbreak. She is free to explore the world and interact with other characters while forming questions, generating hypotheses, collecting data, and testing her hypotheses. Throughout the mystery, she can walk around the island and visit the infirmary, the lab, the dining hall, and the living quarters of each member of the team. She can pick up and manipulate objects, and she can talk with characters to gather clues about the source of the disease. In the course of her adventure she must gather enough evidence to correctly identify the type and source of the disease that has infected the camp members.

Affective support in CRYSTAL ISLAND attempts to mirror the appraisal process by considering student and agent goals and how these correspond to events in the environment (Figure 2). Each student is assumed to have their own goals for their interaction with the environment, such as learning as much about microbiology or being the first to solve the mystery. Their emotional states are then based on cognitive appraisals of how these goals are being met based on their current state in the environment. These states result in action tendencies and coping strategies that may manifest in the student’s behavior in the virtual environment. For example, students may respond to negative emotion with problem-focused strategies, such as working more diligently to solve the mystery, or with emotion-focused strategies such as disengagement from the task.
Alternatively, empathetic agents in the environment have their own internal goals and beliefs that are focused on producing optimal learning and affective experiences for the student. In order to achieve this goal the agent first must assess the student’s emotions and behavior patterns in the environment by simulating an appraisal process on the part of the student. It then compares predicted student affective states and learning behaviors with goals it has for the student. The estimated benefits of certain affective states or behavioral patterns are informed by past empirical evidence on their impact on learning and motivation. Based on this assessment the agent considers multiple strategies of intervening to aid student development. These strategies as well take on a problem-based or emotion-based focus. This distinction mirrors the coping strategies proposed by appraisal theory as well as the distinction between affective and helping behaviors in Davis’ model of empathy. An agent will then deliver a selected feedback strategy based on its expected impact on the student’s motivation and learning.

Figure 2. Model of affective reasoning in CRYSTAL ISLAND
4 Work to Date

Presently, work on affective reasoning in CRYSTAL ISLAND has focused specifically on empirically informing each component of the overall model. While the system has yet to be fully implemented and tested, each segment has been independently validated. A discussion of these results follows.

4.1 Monitored Variables

As noted above, in order to empathize with a human target an empathetic agent can either consider physiological evidence such as facial expressions or context information of the situation. In order to gain the fullest picture of the emotional experience of the student, a variety of variables are monitored during a user’s interaction. These include:

1. **Personal Attributes.** These static attributes help to provide insight into the personality and goals of individual students. These variables are taken directly from students’ scores on personal surveys typically given prior to interaction with CRYSTAL ISLAND. These surveys include measures of goal orientation, personality, empathetic tendencies, and emotion regulation.

2. **Environment Variables.** These dynamic attributes capture a snapshot of the student’s activity in the learning environment at a particular time. They provide a summary of important actions taken, such as TestsRun, BooksViewed, and GoalsCompleted. They also include information about how well the student is doing in the environment based on certain milestones, such as SuccessfulTest and WorksheetChecks. These variables are also used to make assessments about general student behaviors such as how much time the student is spending on tasks that are not central to the learning activity.

3. **Physiological Data.** Physiological data can be used to assess the physical impact of the user’s emotions on their person. Physiological data used in this system have included biofeedback monitors that measure changes in heart rate and galvanic skin response as well as specially-designed seats that measure a student’s posture during their interactions with the environments.

4.2 Simulated Appraisal

After gathering evidence regarding a user’s current situation, the agent must then make an assessment of what they expect the user’s affective state to be at a particular point in time. Recent work has focused on using Bayesian networks for prediction of student affect. The structure of the model (Figure 3) was designed with a focus on the appraisal of learning and performance goals and how these goals were being met based on the status of the game environment as described by Elliot and Pekrun’s model of learning emotions. For example, learning-focused activities such as book reading or note-taking are expected to impact how much a student’s learning goals are
Being met, while performance appraisals are more likely tied to reaching important milestones such as running a successful test. Meanwhile, goal focus and valence tendencies are considered to be dependent on their personal attributes as described by the model. For example, students with approach orientations are expected to have generally more positive temperaments and emotional experiences than students with avoidance orientations. Similarly, personality traits such as agreeableness and openness are expected to contribute to an individual’s overall temperament.

After the structure was designed, the parameters of the model were learned using data from 260 middle school students who interacted with the CRYSTAL ISLAND learning environment [20]. During these interactions students were regularly prompted (every 7 minutes) to provide a self-report of their emotional state from a list of seven states relevant to the learning activity: anxious, bored, confused, curious, excited, focused, and frustrated. The corpus of data was divided into ten independent folds. During each training session, nine folds were used to learn the model parameters and the remaining tenth fold was used for validation. Evaluation of the model showed that the Bayesian network could predict user’s emotional states with 25.5% accuracy and could predict the valence of the emotional state with 66.8% accuracy. Both of these predictions offer a significant gain over the baseline of the most frequent label (22.4% for emotion label and 54.5% for valence) (p<0.05). As an additional comparison, a naïve Bayesian model was learned. A naïve Bayesian network operates under the “naïve” assumption that all variables are directly related to the outcome variable but are conditionally independent of each other [21]. The learned naïve Bayesian network achieved a predictive accuracy of 18.1% on emotion label and 51.2% on valence. These results demonstrate the benefits of using a theoretical model of learner emotions to guide the model’s structure.
However, the simple Bayesian network has no explicit representation of how emotions change over time. For instance, while poor performance at a task may merely be frustrating early in the interaction, for highly performance-oriented students this could turn into anxiety as more and more time passes. In order to capture the dynamic nature of emotions as they occur over time, the structure of the simple Bayesian network was used as the foundation of a series of dynamic Bayesian networks.

Dynamic Bayesian networks extend Bayesian networks by representing changes of the phenomena modeled over time. In this way, observations at time $t_n$ are able to inform observations at time $t_{n+1}$ [21]. A variety of representations of the dynamic nature of appraisal and the resulting affective states were tested. Of these, the model with the highest accuracy was able to predict emotional state with 32.6% accuracy and valence with 72.6% accuracy. This model (Figure 4) included a dynamic link between both emotion and valence, where the values of these two variables at $t_{n+1}$ are partially informed by the emotion and valence at time $t_n$. Together these models show the benefit gained from using theoretical models of learner emotions to predict student affect as well as the importance of considering the dynamic nature of emotional experience.

4.3 Estimated Utilities

After the agent has made a prediction of the emotional state it believes the student to be experiencing it must then assess how this state compares to the agent’s own goals for the student. Overall the agent values learning, positive affect, and focused engagement within the environment. Empirical findings relating emotional states and behaviors to learning outcomes and motivation provide guidance for the agent when assessing how well the student’s current states align with the agent’s own goals.
For instance, from the corpus of 260 students previously described, strong evidence was found correlating positive emotion with learning gains between a pre- and post-interaction content test. Positive affect was strongly correlated with learning gains, \( r(258) = 0.176, p = 0.004 \), while negative affect was negatively correlated with learning gains, \( r(258) = -0.147, p = 0.017 \). These results corroborate with many findings in the psychological community which suggest that positive affect can lead to increased learning through better strategy selection [22,23], increased persistence [24], and improved use of mental resources [25]. Alternatively, negative emotions are believed to lead to decreased motivation and effort [23,26] and a desire to avoid the task [27].

This tendency is further evidenced when examining student’s off-task behavior within CRYSTAL ISLAND. As a highly interactive game-based learning environment students are often distracted by the game features and engage in off-task behaviors such as climbing on buildings or experimenting with in-game physics and objects [28]. However, interesting relationships have been found between student’s affect and their likelihood to disengage from the learning task and focus on game-centric features of the learning environment. For instance, the emotional state of boredom was highly correlated with disengaged, off-task behavior, \( r(258) = 0.226, p < 0.001 \), while curiosity was negatively correlated with off-task behavior, \( r(258) = -0.138, p = 0.031 \). In general, the tendency to disengage and go off-task is also negatively correlated with a student’s overall learning gains, \( r(258) = -0.167, p = 0.009 \), but there is evidence that some students may be using off-task behavior as an effective way to self-regulate their emotional states [28]. These students disengage in times of frustration and confusion but return shortly thereafter refreshed and ready to tackle the learning task. Further distinguishing the beneficial and harmful effects of off-task behavior is an important next step to having full-informed estimates of utility for specific affective states and behavior tendencies.

**4.4 Coping Action Selection**

There are a variety of ways in which an agent can respond once it has assessed the student’s emotional state and how this state relates to the agent’s goals of student learning. First, the agent may have several types of affective responses. They may experience and express the same emotional state they attribute to the user (parallel empathy) or they may have an affective response of an emotional state that is reactive to that of the student (reactive empathy) [6]. For example, if a student is frustrated, a parallel empathetic response would include a mimicked sense of frustration, while a reactive response may range from hope and encouragement for the student to continue, or alternatively anger and disappointment if the student disengages from the environment as a result of this frustration (e.g., [29]). The agent may also have a cognitive response in which they do not express any emotional state but may still provide helpful hints and suggestions to students to help them maintain a positive state or overcome the causes of a negative state.

Determining which type of empathetic response is most beneficial has been a focus of recent work. First, machine learning techniques were utilized to distinguish
whether parallel or reactive affect-focused empathy would be most beneficial in specific cases [30]. Detailed investigation of student responses to empathetic agent feedback indicated some interesting differences in how students react to exhibitions of parallel and reactive empathy. Students met with parallel empathy had a strong tendency to remain in the same or similar affective state [30]. This was true whether the student was in a positive or negative state. Alternatively, students who received reactive responses tended to transition to very different states. If the student was exhibiting a negative state, a reactive response would successfully encourage them to enter a more positive affective state. Unfortunately, reactive responses would bring students in a positive state down to a more negative state, often confusion. It appears that attempting to further motivate a student who is already feeling positive has unfortunate side effects.

Subsequent work built upon these results to further identify cases where cognitive empathetic responses, or helping behaviors, were more beneficial when compared to affective empathetic response [31]. This work asked students to rate the helpfulness of feedback statements which were randomly selected between empathetic responses and task-based hints. Incorporating previous findings, students received a parallel empathetic statement in response to positive emotion and reactive responses to negative affective states. For task-based feedback, students would receive a summary of current progress when they reported a positive emotional state and would be given additional hints to help them overcome a negative state if it was reported. In general, task-based feedback was rated more helpful by students than the empathetic responses. While in most cases the ratings for task-based feedback were significantly higher than empathetic responses (p < 0.001), this was not the case for the emotions that were not directly associated with learning, (e.g. excitement). These emotions, which tie more closely with the narrative plot of the environment, were best met with empathetic responses.

Along these lines, recent work has indicated that the selection of an appropriate empathetic response is very important in maintaining a positive user experience [32]. Analyses showed that empathetic responses that were deemed inappropriate by the user often had negative consequences on the user’s affective state. In this case the agent was making an inappropriate action selection with knowledge of the student’s state through self-report. Unfortunately this issue is compounded by the imperfect performance of predictive models of student affect. Empathetic agents may make incorrect decisions both at the time of diagnosis and intervention resulting in undesirable consequences. This highlights the importance of incorporating notions of confidence and risk when the agent is making decisions about empathetic behaviors. It is hoped that an agent capable of weighing its confidence in its diagnosis and selection versus the risk in case of an error will be the most effective at encouraging positive affect and learning for students.

5 Discussion

This work presents an appraisal-based framework for empathetic pedagogical agents. Validation of each component of the system shows significant promise. While the
results of the investigations are encouraging, it is important to note some important limitations. First, the studies rely heavily on user feedback, i.e., users are asked to report both the emotion they are feeling at any given time and their perception of agent behavior. Self-reports such as these are problematic, in that users may not be able to accurately distinguish their emotional state or may choose not to report it correctly. Consequently, replicating these analyses with expert-judged or physiologically based assessments of emotion could yield a more accurate view of students’ true emotions. Additionally, this work focuses only on a small subset of theoretical models that have been proposed for affect and empathy. In order to further validate theoretical models and their usefulness for informing virtual agent behavior in learning environments it will be important to consider other existing theoretical models. It will also be important to determine how well-suited this framework is to other learning environments, both traditional computer-supported tutoring systems as well as game-based learning environments.

These limitations point to several important areas of future work. The next proposed step is to determine what level of accuracy is needed for predictive models of affect to inform affective feedback and to investigate how action selection methods can be optimized to improve student affect and learning. The integration of prediction and decision making is an important final step in completing the proposed model of agent affect and empathy. Once accomplished, the entire framework, rather than each component, can be validated as an appropriate mechanism for modeling empathetic behavior in a virtual learning environment.

Additionally, future work is needed to determine whether the proposed model performs well in other intelligent learning environments. CRYSTAL ISLAND differs from most learning scenarios in that it is an open-ended exploratory environment without a clearly defined problem space and its affective models may not generalize to traditional learning environments. Additionally, it will be interesting to explore other theoretical models of learner emotions to compare how well these translate into computational models. This line of work may also help to validate theoretical models of learner emotions. Finally, a more comprehensive set of learning-focused cognitive and affective states could provide additional power to affect-sensitive systems. For example, the current predictive models do not consider a “neutral” affective state, nor do they offer a clear distinction between traditionally cognitive states (e.g. focused) and affective states (e.g. excited). Representations that distinguish these states may improve predictive and responsive capabilities.

6 Conclusion

This work presented a model of empathy for virtual pedagogical agents based on several psychologically grounded models of affect and empathy. This technique was selected because no uniformly accepted model of either phenomenon exists. Instead there are a variety of psychological models, many of which are particularly tailored to specific applications or domains (such as learning). By selecting models that are relevant to the application domain (virtual agents in a game-based learning environment), we are able to capitalize on the specific benefits of each. Additionally,
by separately validating each component we are able to ensure the selected psychological models and their computational interpretations are appropriate. However, an important area of future work is to completely integrate each component of the model and evaluate its effectiveness at empathizing with student users. In this way, the entire framework, rather than each component, can be validated as an appropriate mechanism for modeling empathetic behavior in a virtual learning environment.

Acknowledgments. The authors wish to thank members of the IntelliMedia Group for their assistance, Omer Sturlovich and Pavel Turzo for use of their 3D model libraries, and Valve Software for access to the Source™ engine and SDK. This research was supported by the National Science Foundation under Grants REC-0632450, DRL-0822200, IIS-0812291, and CNS-0739216. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship.

References
