Learning Empathy: A Data-Driven Framework for Modeling Empathetic Companion Agents

Scott W. McQuiggan
Department of Computer Science
North Carolina State University
Raleigh, NC 27695, USA
swmcquig@ncsu.edu

James C. Lester
Department of Computer Science
North Carolina State University
Raleigh, NC 27695, USA
lester@csc.ncsu.edu

ABSTRACT
Affective reasoning plays an increasingly important role in cognitive accounts of social interaction. Humans continuously assess one another's situational context, modify their own affective state accordingly, and then respond to these outcomes by expressing empathetic behaviors. Synthetic agents serving as companions should respond similarly. However, empathetic reasoning is riddled with the complexities stemming from the myriad factors bearing upon situational assessment. A key challenge posed by affective reasoning in synthetic agents is devising empirically informed models of empathy that accurately respond in social situations. This paper presents CARE, a data-driven affective architecture and methodology for learning models of empathy by observing human-human social interactions. First, in CARE training sessions, one trainer directs synthetic agents to perform a sequence of tasks while another trainer manipulates companion agents' affective states to produce empathetic behaviors (spoken language, gesture, posture). CARE tracks situational data including locational, intentional, and temporal information to induce a model of empathy. At runtime, CARE uses the model of empathy to drive situation-appropriate empathetic behaviors. CARE has been used in a virtual environment testbed, and an evaluation suggests that the CARE paradigm can provide the basis for effective empathetic behavior control for companion agents.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information Systems**]: Artificial, augmented, and virtual realities; Evaluation/methodology.

General Terms
Algorithms, Design, Experimentation, Human Factors.

Keywords
Synthetic Agents, Affective Reasoning, Empathy.

AAMAS'06, May 8-12, 2006, Hakodate, Hokkaido, Japan.

1. INTRODUCTION
Recent years have witnessed significant progress on synthetic agents. With a broad range of applications in entertainment, education, and training, foundational work on synthetic agents has yielded expressive models of embodied cognition and behavior that support rich interactions in virtual environments [1, 2, 6, 16, 20, 27]. Complementing advances in cognition and behavior, affective reasoning [10, 12, 22, 25] has begun to play a central role in synthetic agents [4, 5, 20, 21], and the community is now well positioned to investigate affective reasoning in the context of social interaction [17, 23, 26]. Transitioning affective synthetic agents into the social arena could yield companion agents that provide motivating support and comfort. *Companion agents* can facilitate social interaction, a critical capability in virtual environments for education [5, 7, 20] and training [26]. Companion agents help users cope with frustration [5], deal with stress [26], and counsel children on social behaviors [23].

Empathy is a key component of social interaction [14]. Because empathetic companion agents hold much promise for socially engaging virtual environments, empathy modeling is a logical next step in the evolution of synthetic agents. One can distinguish two fundamental approaches to modeling empathy: analytical and empirical. In the *analytical* approach, models of empathy can be constructed by analyzing the findings of the empathy literature. However, empathy is not well understood. It is only in the past two decades—this is very recent in the history of psychology—that empathy has become a focus of study for social psychologists [9]. Perhaps as a result of its limited study, while we have expressive computational models of affect, e.g., the OCC model [22], we do not have similarly rich models of empathy. Moreover, because empathetic reasoning requires drawing inferences about another's intentions, her affective state, and her situational context, devising a universal model of empathy seems to be well beyond our grasp at the current juncture.

An alternative to analytically devising models of empathy for synthetic agents is the *empirical* approach. If somehow we could create models of empathy that were derived directly from observations of "empathy in action," we could create empirically grounded models based on human-human empathetic behaviors exhibited during the performance of a specific task within a given domain. While it is not apparent that this approach could produce a universal model of empathy—a universal model may not even be achievable, at least in the near term—the empirical approach could nonetheless generate models of empathy that significantly

Figure 1: Treasure Hunt virtual world with companion agent (left) and the target's agent (right).

extend the communicative capabilities of socially intelligent companion agents.

The empirical approach calls for a data-driven framework for modeling empathy. This paper presents CARE,[1] a data-driven affective architecture and methodology for learning empirically informed models of empathy from observations of human-human social interactions. During training sessions, CARE monitors situational data including locational, intentional, and temporal information while one trainer (the *target*) directs her synthetic agent to perform a sequence of tasks in a virtual environment as another trainer (the *empathizer*) reactively manipulates her synthetic agent's affective state to produce empathetic behaviors (spoken language, gesture, and posture). Inducing a model of empathy, CARE uses situational data as predictive features for empathy assessment (when to exhibit an empathetic behavior) and for empathy interpretation (which levels of valence and arousal should be chosen, i.e., the affective state). At runtime, CARE uses the resulting model to drive situation-appropriate empathetic behaviors in the companion agent as it interacts with actual users.

The paper is organized as follows: Section 2 provides background on empathy and affective reasoning in synthetic agents. Section 3 presents the CARE architecture and methodology. CARE has been used to create a model of empathy for a companion agent inhabiting Treasure Hunt (Figure 1), a virtual environment in which a user and a companion agent search for treasures. Section 4 describes the CARE implementation and its generation of the empathy models for the Treasure Hunt companion agent. Section 5 presents an evaluation of CARE based on a 31-subject experiment. The study suggests that CARE models can provide the basis for effective empathetic behavior control for companion agents.

2. Empathetic Synthetic Agents

Devising computational models of empathy contributes to the broader enterprise of modeling affective reasoning [24]. Beginning with Elliott's implementation [10] of the OCC model [22], advances in affective reasoning have accelerated in the past few years, including the appearance of a sophisticated theory of appraisal [12] based on the Smith and Lazarus Appraisal Theory [19]. We have also begun to see probabilistic approaches to assessing users' affective state in educational games [7] and investigations of the role of affect and social factors in pedagogical agents [3, 5, 11, 17, 20, 26]. Recent work on empathy in synthetic agents has explored their affective responsiveness to biofeedback information and the communicative context [26]. It has also yielded agents that interact with one another and with the user in a virtual learning environment to elicit empathetic behaviors from its users [23].

Empathy is a complex socio-psychological construct. Defined as "the cognitive awareness of another person's internal states, that is, his thoughts, feelings, perceptions, and intentions" [15], empathy enables us to vicariously respond to another via "psychological processes that make a person have feelings that are more congruent with another's situation than with his own situation" [14].

Social psychologists describe three constituents of empathy. First, the *antecedent* consists of the empathizer's consideration of herself, the target's intent and affective state, and the situation at hand. Second, a*ssessment* consists of evaluating the antecedent. Third, *empathetic outcomes*, e.g., behaviors expressing concern, are the products of assessment [9] including both affective and non-affective outcomes (e.g., judgment, cognitive awareness). Two types of affective outcomes are possible. In *parallel outcomes*, the empathizer mimics the affective state of the target. For example, the empathizer may become fearful when assessing a target's situation in which the target is afraid. In *reactive outcomes*, empathizers exhibit a higher cognitive awareness of the situation to react with empathetic behaviors that do not necessarily match those of the target's affective state. For example, empathizers may become frustrated when the target does not meet with success in her task, even if the target herself may not be frustrated. Accurately modeling parallel and reactive empathetic reasoning presents significant challenges.

3. DATA-DRIVEN EMPATHY MODELING

The prospect of creating an "empathy learner" that can induce empirically grounded models of empathy from observations of human-human social interactions holds much appeal. To this end, this paper proposes CARE, an affective data-driven paradigm that learns empathetic assessment (when to be empathetic) and empathetic interpretation (how to be empathetic). CARE consists of a trainable agent architecture and a two-phase methodology of training and learning.

3.1 Architecture

The CARE architecture operates in two modes: empathetic model induction in which the architecture interacts with two trainers (depicted in the diagram with dotted arcs), and runtime operation, in which it manages empathetic behaviors for a companion agent interacting with a user (depicted in the diagram with solid arcs) (Figure 2):

- **Empathetic Model Induction:** Trainers interact with CARE via interfaces through which they direct synthetic agents in the virtual environment. The virtual environment tracks all

[1] CARE: Companion-Assisted Reactive Empathizer.

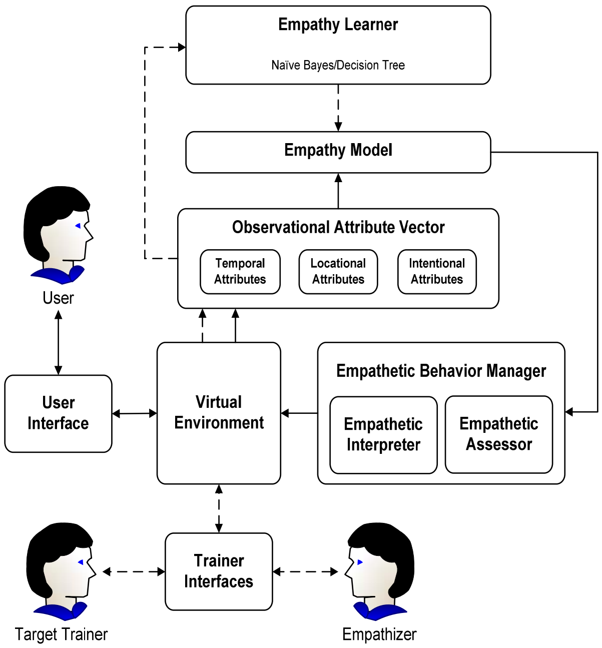

Figure 2: General Architecture

activities in the world and reports observable attributes pertaining to temporal, locational, and intentional information. These are passed to the empathy learner during the training phase. During the subsequent learning phase, the learner induces a model of empathy that is operational, i.e., it can be used at runtime.

- **Runtime Operation:** Users interact with CARE via an interface through which they direct a synthetic agent in the virtual environment. Throughout their experience, they interact with a companion agent controlled by CARE. The virtual environment again tracks all activities in the world and monitors the same observable attributes reported to the empathy learner during empathetic model induction. The induced model is used by the empathetic behavior manager to (1) assess the situation to determine *when* to be empathetic, and (2) interpret situations deemed "empathy-worthy" to decide *how* to be empathetic. When a situation calls for empathy, a suitable empathetic behavior (including speech, gesture, and posture) is selected for execution by the companion agent to react empathetically to the user's situation. Spoken components of companion agent empathetic behaviors explicitly state the affective state being conveyed with commonly associated gestures and posture, i.e., *dropped shoulders*, *arms crossed*, *looking down*, and *head shaking* gestures and postures accompany the companion agent's verbal communication, "This has become quite frustrating."

3.2 Training and Learning

In the training phase, CARE's trainable agent must be exposed to social situations similar to the ones it will encounter at runtime. Because empathy by its very nature involves multiple actors (here we focus on two), the training experience should revolve around

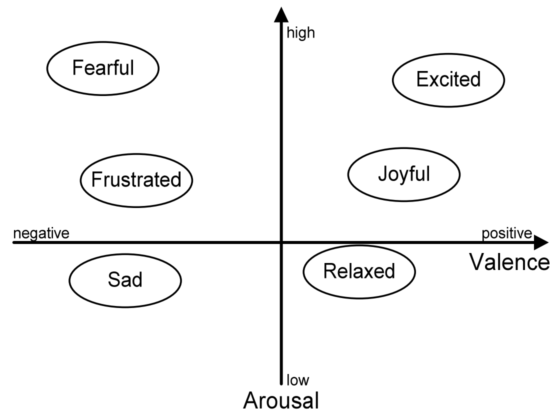

Figure 3: Two-dimensional Affective Space

the interaction of multiple subjects in situations that elicit empathetic behaviors.

CARE training sessions are therefore situated in task-oriented scenarios involving two trainers, a *target* and an *empathizer*, each of whom is represented by a synthetic agent in the 3D virtual environment where training takes place. The target, whom is given a multi-objective mission to complete, controls her agent to navigate and perform tasks in the virtual environment from a first person point-of-view (POV). It is the task of the empathizer, who looks on from a third-person POV, to monitor the target's activities and select suitable empathetic affective states based on the target's observed behaviors. Selecting an affective state causes her agent to perform an empathetic behavior.

To collect empathy data that is as representative as possible of that which will be encountered by the companion agent at runtime, training sessions must satisfy the following requirements:

- ***Affective space coverage***: At each stage of the mission, to promote the target's experiencing a range of emotions spanning the classic two-dimensional affective space defined by *valence* (degree of attraction, ranging from negative to positive) and *arousal* (level of stimulation, ranging from low to high) [18] (Figure 3), the target should be faced with goals of varying degrees of difficulty: some should be very easy to achieve, while others should be very challenging. For example, in Treasure Hunt, the virtual environment that serves as a testbed for CARE, some treasures are in plain view of the target while others are partially occluded and some are hidden altogether. Some targets should be exposed to virtual environments in which goals are easy to achieve, and some should be introduced into worlds in which goals are difficult to achieve. Thus, in some Treasure Hunt worlds, targets can score a specified number of points by collecting treasures very easily, while other worlds pose significant challenges stemming from the accessibility and varying point values. These unique situations offer opportunities for users to experience a variety of reactive emotions.

- ***Double-blind training***: Training sessions should be conducted such that the target is unaware that an empathizer is at the controls of the empathetic behaviors of the companion agent in the virtual world. Likewise, restricting the empathizer's environment to the virtual world (i.e., without access to the target's facial or vocal expressions) enables empathetic decisions to be based solely on inferences from the observed virtual world (thus, similar inferences are likely to be made by the empathy models at runtime).

- ***Empathy-centered control***: The empathizer should be able to focus exclusively on empathy decision making. Thus, navigation control for the companion agent is provided by an autonomous path planning mechanism that ensures that the companion agent is always within a specified proximity to the target's agent in the virtual environment and that the empathizer has an adequate view of the target agent's experiences.

- ***Training session length***: Each training session must strike a careful balance between being long enough to yield a large body of data and short enough so that the trainers do not become overly fatigued. In the Treasure Hunt environment, experimentation indicated that 7-minute sessions satisfied this requirement and provided sufficient data for empathy modeling.

- ***Controlled affective expression***: Minimizing the complexity of the empathizer's task can be achieved by limiting the set of emotions at her disposal. For example, empathizers in Treasure Hunt have access to six affective states: excited, joyful, relaxed, fearful, frustrated and sad. This particular set of emotions was chosen because it covers the four quadrants of the two-dimensional affective space [18] and considers three levels arousal (high, medium and low) for each level of valence (positive or negative).

- ***Uniform agent personae***: While investigating different personae is a promising direction for future work, e.g., pedagogical agent personae experiments [3], baseline training should control for personae by holding both the target's agent and the empathizer's agent constant throughout training sessions.

- ***Situation data collection intervals***: Situation data should be collected at least as often as significant events occur, where an event is deemed significant if it can plausibly affect the empathizer's decisions. In Treasure Hunt, locational data were collected when events in the world indicated notable state changes, e.g., an agent's entering of a room, while some temporal data were monitored continuously, e.g., the amount of time between goal achievement. A typical training session in Treasure Hunt yields approximately 6,000-9,000 data points.

Accurately modeling empathy requires a representation of the situational context that satisfies two requirements. First, it must be sufficiently rich to support empathetic assessment and empathetic interpretation. Second, it must be encoded with features that are readily observable at runtime so that they may drive companion agents' empathetic decision making. CARE therefore employs an expressive representation of all activities in the virtual environment by encoding them in an observational attribute vector that is used in both modes of operation: during empathetic model induction, the *observational attribute vector* is passed to the empathy learner for model generation; during runtime operation, the attribute vector is monitored by the empathetic behavior manager for determining empathetic

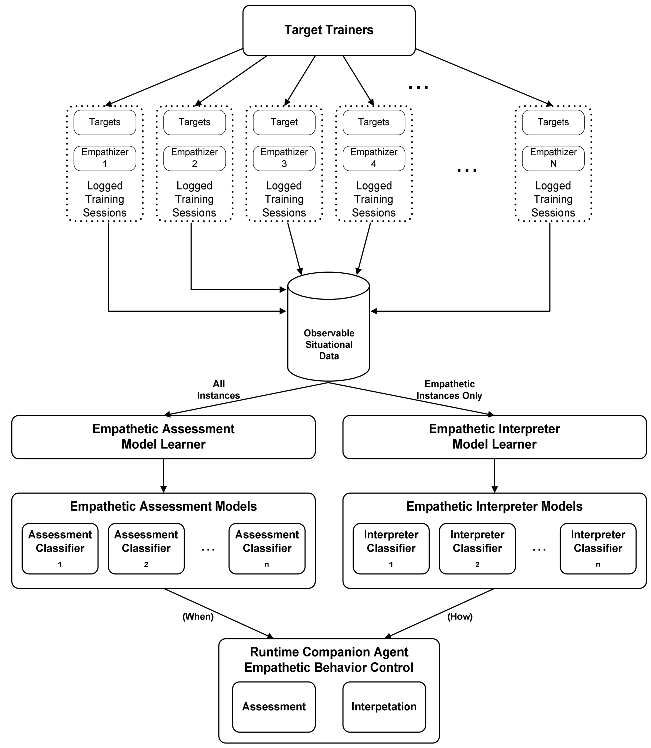

Figure 4: CARE Framework Data Flow.

behavior. CARE's observable attribute vector represents three interrelated categories of features for making empathetic decisions:

- **Temporal features**: CARE tracks the amount of time that has elapsed since the target/user arrived at the current location, since the target/user achieved a goal, since the empathizer/companion agent last behaved empathetically, and since the target/user was last presented with an opportunity to achieve a goal.

- **Locational features**: CARE continuously tracks the location of all agents in the environment. It monitors locations visited in the past, locations recently visited, locations not visited, and locations being approached. Locations are associated with specific areas in the virtual environment or areas containing significant objects or obstacles, e.g., goals or locked doors.

- **Intentional features**. CARE tracks goals being attempted (as inferred from locational and temporal features, e.g., approaching a location where a goal can be achieved), quantity and quality of goals achieved, the rate of goal achievement, and the effort expended to achieve a goal (as inferred from recent exploratory activities and locational features).

In the CARE implementation for Treasure Hunt, the observational attribute vector encodes 192 features. During empathetic model induction, an instance of the vector is logged every time a significant event occurs. On average, vectors are updated several hundred times each minute. At runtime, the same features are updated continuously by the virtual environment and are used by the empathetic behavior manager to select situation-appropriate empathetic behaviors. Figure 4 shows how information from

Figure 5: A relaxed companion agent and target agent.

Figure 6: A frustrated companion agent and target agent.

observations in CARE training sessions flows from the training phase to the learning phase for empathetic model induction.

Finally, in the learning phase, CARE induces a dual model of empathy. One component will be used at runtime to support empathetic assessment, and the other will be used to support empathetic interpretation. CARE's empathy learner first uses all of the data collected in the training session to induce the empathetic assessment model. Induction may be based on any standard classifier learning technique.

Two versions of CARE have been implemented in Treasure Hunt, one with a naïve Bayes classifier and one with a decision tree classifier. The evaluation reported in Section 5 discusses the performance of both approaches. CARE's empathy learner next uses a subset of the data collected in the training sessions to induce the empathetic interpretation model. Here, it only considers data instances in which empathy was in fact exhibited. The second induction produces a model of empathy interpretation that at runtime is used to guide agent's empathetic behaviors. The products of the learning phase are two classifiers used to determine when and how the companion agent should be empathetic as dictated by a generalized model induced from all of the empathizing trainers' empathetic behavior decisions. Because the classifiers employ features directly observable in the environment, they can be easily integrated into the runtime behavior control systems of companion agents in the form of rules or probabilistic statements.

4. EXAMPLE APPLICATION

The CARE paradigm has been used to train models of empathy and to control the behavior of a companion agent at runtime in Treasure Hunt, a virtual environment testbed in which targets/users are instructed to collect treasures in the allotted time. After introducing the Treasure Hunt virtual environment, we describe the implementation and present an illustrative example of CARE'S generating an empathetic behavior.

4.1 Treasure Hunt Virtual Environment

Treasure Hunt is a prototype virtual environment featuring a synthetic agent controlled by the user and a companion agent whose empathetic behaviors are controlled by CARE. The user navigates the 3D virtual world in search of hidden (and some not-so-hidden) treasures. Each treasure box is labeled with the value of its contents, representing points the user obtains when collecting the associated treasure. Throughout the users' quest for treasure, the companion agent follows along and expresses empathetic behaviors as appropriate situations arise in the users' experiences (Figure 5).

4.2 Implementation

CARE's empathetic assessment model and interpretation model have been implemented using naïve Bayes and decision tree approaches. A discussion of their relative performance follows in Section 5. The empathetic models were induced from a dataset consisting of a 192-dimensional observational attribute vector. The observational attribute vector consists of temporal, locational, and intentional features. For example, a sliding ten-second window was used as a temporal feature for tracking user goal attainment, while a binary locational feature monitored whether the user had yet visited the docks or rocky beach area, and an intentional feature was used to detect when the user was moving in the direction of a high-valued goal in the her view.

Treasure Hunt was implemented using a high-performance 3D game platform from Valve Software. The virtual environment, observational attribute monitoring, and empathetic models were implemented with Valve's Source™ engine (the game engine for Half-Life[2]) and the accompanying 3D game platform for Half-Life[2].

4.3 Example Scenario

To illustrate the empathetic behavior control posed by CARE, consider the following scenario, which repeatedly played out in CARE training sessions. As we catch up with the user, she has navigated her synthetic agent throughout the virtual environment struggling to find significant, high-valued treasure. The user and empathizer are aware that the user has not yet met her expected treasure collection quota (as depicted in the graphical HUD representation in the bottom corner of the display) and is quickly running out of time. Only 30 seconds remain.

Now, the user has found her agent's way into a location on the beach of the Treasure Hunt virtual environment, a location visited by the user's agent in the early moments when the session began. The empathizer realizes that this particular location has been previously visited and was already determined to be an area without any treasure boxes. It has now been over one minute since the user last discovered any treasure at all. Assessing the situation, the empathizer selects the frustrated affective state, thereby initiating a behavioral sequence in which the companion agent announces her frustration by directly stating, "This is becoming quite frustrating," and using gestures and posture similar to the companion agent depicted in Figure 6. (The agent's speech segments are stored in pre-rendered audio clips.)

CARE's empathy learner monitored a variety of environmental characteristics, including those described above, during its training sessions. These recorded instances aid the empathy models in reproducing similar appropriate inferences in analogous situations where time is running out, the user's agent is in a previously visited location known to be without treasure, and the user's intended treasure collection goal is likely to fail. Thus, given the same situation with CARE driving the empathetic behaviors of the companion agent at runtime, empathetic assessment and interpreter models are likely to make similar appropriate empathetic decisions. The following section discusses how accurately the models learned by the agent are able to predict empathizer actions.

5. EVALUATION AND DISCUSSION

This section discusses the user study and CARE training sessions, assesses the naïve Bayes and decision tree classification approaches used for modeling empathetic assessment (when) and empathetic interpretation (how), and suggests several design implications.

5.1 User Study

In a formal evaluation, more than two hours of data were gathered from thirty-one subjects in an Institutional Review Board (IRB) of North Carolina State University approved user study. Participation included 25 targets and 6 empathizers. There were 20 male subjects and 5 female subjects varying in race, ethnicity, age and martial status who participated as training targets. There were 3 male and 3 female subjects participating as training empathizers. On average, empathizers completed 4 training sessions (SD = 1.94), each with its own unique target trainer.

Before empathizers began training, they completed Davis's Interpersonal Reactivity Index (IRI) [8], which provides a measure of an individual's empathetic nature. The index consists of 28 statements in which respondents are instructed to rate the degree to which each item describes them on a Likert scale of 0 to 4. The result is a set of 4 subscale values quantifying the following qualities of empathy: fantasy, perspective taking, empathic concern and personal distress [10]. Female empathizers scored higher than male empathizers in each quality except for perspective taking, in which the males scored one-half point higher than the female empathizers. Subjects were found to be representative of the general population in empathetic qualities, as reported in [8].

Next, empathizers were provided standard definitions of emotion and empathy, similar to those discussed in Section 2, and they were instructed to rely on these definitions for their empathetic decision making. In the meantime, training targets were prepared for the session and completed a pre-session demographic survey. Training targets were unaware of the empathizer's role until the training session was complete. In fact, the empathizer and target were housed in separate rooms for the entire session and were never exposed to the other's presence.

Once sessions began, empathizers observed training target interactions and made empathetic behavioral decisions based on the situation with which they were presented. Sessions lasted for 7 minutes. Both training targets and empathizers received post-session surveys and were interviewed to gain insight into the overall experience. In these interviews it was often mentioned by training targets that they felt the role of the companion agent was solely to provide "emotional commentary." In post-interviews with empathizers it was discovered that up to an additional set of 4 emotions, for a set of set of 10 emotions, would have been preferred.

5.2 Experiment

Both naïve Bayes and decision tree models were induced from data collected in the training sessions described above. As noted earlier, 192 observational attributes were used to define the feature vectors. Naïve Bayes and decision tree classifiers are effective machine learning techniques for generating preliminary predictive models. Naïve Bayes classification approaches produce probability tables that can be incorporated into runtime systems and used to continually update probabilities for monitoring when and how to be empathetic. Decision trees provide interpretable rules that support runtime decision making. Both the naïve Bayes and decision tree machine learning classification techniques are useful for preliminary predictive model induction for large multidimensional data. Because it is unclear precisely which runtime variables are likely to be the most predictive, naïve Bayes and decision tree modeling provide useful analyses that can inform more expressive machine learning techniques (e.g., Bayesian networks) that also leverage domain experts' knowledge.

All models were evaluated using a tenfold cross-validation scheme for producing training and testing datasets. In this scheme, data is decomposed into ten equal partitions, nine of which are used for training and one used for testing. The equal parts are swapped between training and testing sets until each partition has been used for both training and testing. Tenfold cross-validation is widely used for obtaining a sufficient estimate of error [28].

5.3 Results

Cross-validated ROC curves are useful for presenting the performance of classification algorithms for two reasons. First, they represent the positive classifications (true positives), included in a sample, as a percentage of the total number of positive classifications along the vertical axis, against the negative classifications (false positives) as a percentage of the total number of negative classifications [28]. Second, the area under ROC curves is widely accepted as a generalization of the measure of the probability of correctly classifying an instance [13]. Figure 7 shows the ROC curves for CARE's naïve Bayes and decision tree classification approaches for modeling empathetic assessment. Figure 8 shows ROC curves for CARE's naïve Bayes and decision tree classification approaches for empathetic interpreter modeling. Associated areas under each of the curves can be found in the figures' captions.

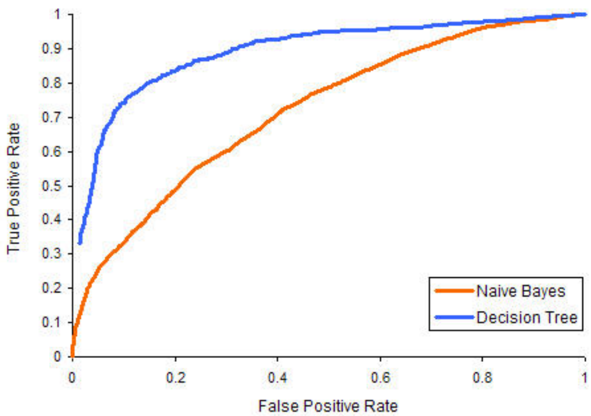

Figure 7: Empathetic Assessment (When) ROC Curves. The ROC curves for each model predicting empathetic behavior triggers in a ten second interval. The area under the naïve Bayes curve is 0.72 and the area under the Decision Tree curve is 0.89.

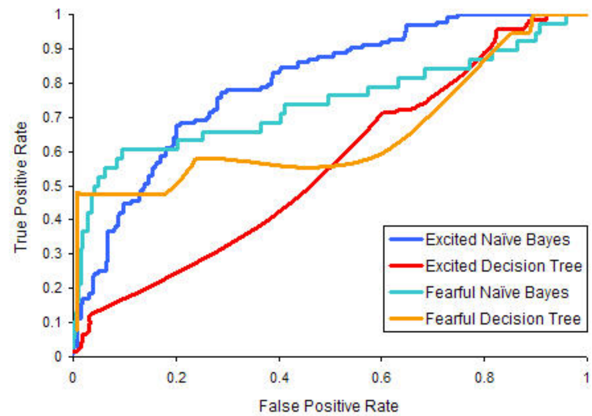

Figure 8: Empathetic Interpretation (How) ROC Curves. The ROC curves represent naïve Bayes and decision tree models for empathetic interpretation for the affective states *excited* and *fearful*. Areas under each curve are as follows: 0.80 (Excited Naïve Bayes), 0.56 (Excited Decision Tree), 0.74 (Fearful Naïve Bayes), and 0.66 (Fearful Decision Tree).

Two categories of functionality can be distinguished. First, the decision tree classifier was best suited for modeling empathy assessment, i.e., it was better able to determine *when* to be empathetic (Figure 7). Second, the naïve Bayes classifier was best suited to modeling empathy interpretation, i.e., it was better able to determine *how* to be empathetic (Figure 8). Although the figure shows only the excited and fearful emotions, all six emotions were evaluated and the naïve Bayes classifier bested the decision tree classifier in every case.

The smoothness of the curve in Figure 7 indicates that sufficient data seems to have been used for training empathy assessment, while the jaggedness of the curve in Figure 8 indicates that more data covering a large space of situations is called for in training empathy interpretation. For example, many empathizers only rarely used particular emotions, e.g., sad, and some trainers suggested that having more affective states available would have been helpful. In general, however, it appears that effective classifiers can indeed be learned for both empathy assessment and empathy interpretation.

6. CONCLUSIONS AND FUTURE WORK

Recent advances in affective reasoning have demonstrated that emotion plays a central role in human cognition and should therefore play an equally important role in synthetic agents. A key affective capability of human social intelligence is empathy. Because empathy is paramount in successful human-human interactions, it would therefore be useful to endow companion agents who are to accompany users in interactive virtual environments with the ability to empathize. Empathy modeling requires accurately assessing a social situation context in order to determine (1) if an empathetic reaction is warranted, and (2) if so, what sort of empathetic behavior should be performed.

This paper has presented a data-driven approach to learning empirically grounded models of empathy from observations of human-human social interactions. In this approach, training data is first generated as a by-product of trainers' interactions with a virtual environment, and models of empathy are induced from the resulting data sets. Critically, the training data employs only observable features, i.e., features that can be directly observed in the environment, so that at runtime, the same features can be used by the empathy models to drive the behavior of companion agents interacting with users. An evaluation of an implemented data-driven empathy modeler suggests that this empirical paradigm offers a promising technique for extending the affective capabilities of synthetic agents. Coupling models of social constructs with expressive controls of agent behavior could perhaps contribute to a new generation of socially and emotionally intelligent synthetic agents in the coming years.

In the future, it will be important to investigate mechanisms for varying empathetic responses in a manner that is most appropriate for individual users, perhaps integrating them with tools such as socio-psychologically validated empathy response instruments. It will also be important to devise integrated methods for employing user physiological response monitored via biofeedback apparati with empirically grounded models of empathy to further extend their range and increase their accuracy. Agent personae offer another direction for future research. For example, females tend to be more empathetic, supported by the fact that a disproportionately large percentage of all empathetic instances in CARE training sessions were performed by female trainers (one-half of all training empathizers were female). Furthermore, exploring models of empathy induced from attributes monitored at varying levels of abstraction may yield models that are transferable between different environments.

7. ACKNOWLEDGMENTS

The authors would like to thank members of the IntelliMedia Center for Intelligent Systems at North Carolina State University for their contributions to the implementation of CARE. Special thanks to Seung Lee for assisting with data analysis. The authors

also wish to thank Valve Software for authorizing the use of their Source™ engine and SDK.

8. REFERENCES

[1] André, E., and Müller, M. E. Learning affective behavior. In *Proceedings of the 10th International Conference on Human-Computer Interaction* (Heraklion, Crete, Greece, June 22-27, 2003). Lawrence Erlbaum, Mahwah, NJ, 2003, 512-516.

[2] Bates, J. *The role of emotion in believable agents.* Technical Report CMU-CS-94-136, Carnegie Mellon University, Pittsburgh, PA, 1994.

[3] Baylor, A. L. The impact of pedagogical agent image on affective outcomes. In *Proceedings of the Workshop on Affective Interactions: Computers in the Loop, International Conference on Intelligent User Interfaces* (San Diego, CA January 9-12, 2005).

[4] Bickmore, T. *Relational Agents: Effecting Change through Human-Computer Relationships*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2003.

[5] Burleson, W., and Picard, R. W. Affective agents: sustaining motivation to learn through failure and a state of stuck. In *Proceedings of Workshop of Social and Emotional Intelligence in Learning Environments*, in *conjunction with the 7th International Conference on Intelligent Tutoring Systems* (Maceio, Alagoas, Brazil, August 30 – September 3, 2004).

[6] Cavazza, M., Charles, F., and Mead, S. J. Interacting with virtual characters in interactive storytelling. In *Proceedings of the 1st International Conference on Autonomous Agents and Multi-Agent Systems* (Bolonga, Italy, July 15-19, 2002). ACM Press, New York, NY, 2002. 318-325.

[7] Conati, C. Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, 16:555-575, 2002.

[8] Davis, M. H. Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126, 1983.

[9] Davis, M. H. *Empathy: A Social Psychological Approach*. Brown and Benchmark Publishers, Madison, WI, 1994.

[10] Elliott, C. *The Affective Reasoner: A Process Model of Emotions in a Multi-agent System*. Ph.D. Thesis, Northwestern University, Chicago, IL, 1992.

[11] Elliott, C., Rickel, J., and Lester, J. Lifelike pedagogical agents and affective computing: an exploratory synthesis. *Artificial Intelligence Today*, Lecture Notes in Artificial Intelligence (Sub series of LNCS), Special Volume 1600, Wooldridge, M. and Veloso, M. (eds.), 195-212, Springer-Verlag, Berlin, 1999.

[12] Gratch, J., and Marsella, S. A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research*, 5(4):269-306, 2004.

[13] Hanley, J.A., and McNeil, B.J. The meaning and use of the area under the receiver operating characteristic (roc) curve. *Radiology* 143:29-36, 1982.

[14] Hoffman, M. L. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge University Press, Cambridge, UK, 2000.

[15] Ickes, W. *Empathic Accuracy*. New York: Guilford, 1997.

[16] Johnson, W. L., and Rickel, J. Steve: an animated pedagogical agent for procedural training in virtual environments. *SIGART Bulletin* 8:16-21, 1998.

[17] Johnson, W. L., and Rizzo, P. Politeness in tutoring dialogs: "run the factory, that's what I'd do". In *Proceedings of the 7th International Conference on Intelligent Tutoring Systems* *(*(Maceio, Alagoas, Brazil, August 30 – September 3, 2004). Springer-Verlag, New York, NY, 2004, 67-76.

[18] Lang, P. J. The emotion probe: studies of motivation and attention. *American Psychologist*, 50(5):372-285, 1995.

[19] Lazarus, R. *Emotion and Adaptation*. Oxford University Press, UK, 1991.

[20] Lester, J. C., Towns, S. G., Callaway, C. B., Voerman, J. L., and FitzGerald, P. J. Deictic and emotive communication in animated pedagogical agents. In *Embodied Conversational Agents*, Cassell, J., Prevost, S., Sullivan, J., and Churchill, E. (Eds.), pp. 123-154, MIT Press, Cambridge, 2000.

[21] Marsella, S., and Gratch, J. Modeling coping behavior in virtual humans: don't worry, be happy. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems* (Melbourne, Australia, July 14-18, 2003). ACM Press, New York, NY, 2003. 313-320.

[22] Ortony, A., Clore, G. L., and Collins, A. *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, MA, 1988.

[23] Paiva, A., Dias, J., Sobral, D., Aylett, R., Woods, S., Hall, L., and Zoll, C. Learning by feeling: evoking empathy with synthetic characters. *Applied Artificial Intelligence*, 19:235-266, 2005.

[24] Picard, R. *Affective Computing*. MIT Press, Cambridge, MA, 1997.

[25] Porayska-Pomsta, K., and Pain, H. Providing cognitive and affective scaffolding through teaching strategies: applying linguistic politeness to the educational context. In *Proceedings of the 7th International Conference on Intelligent Tutoring Systems (*(Maceio, Alagoas, Brazil, August 30 – September 3, 2004). Springer-Verlag, New York, NY, 2004, 77-86.

[26] Prendinger, H., and Ishizuka, M. The empathic companion: a character-based interface that addresses users' affective states. *Applied Artificial Intelligence*. 19:267-285, 2005.

[27] Swartout, W., Gratch, J., Hill Jr., R. W., Hovy, E., Lindheim, R., Marsella, S., Rickel, J., and Traum, D. Simulation meets Hollywood: integrating graphics, sound, story and character for immersive simulation. In Stock, O., and Zancznaro, M. (eds.), *Multimodal Intelligent Information Presentation*, Kluwer, 2004.

[28] Witten, I., and Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd Edition, Morgan Kaufman, San Francisco, CA, 2005.