Available online at www.sciencedirect.com
ScienceDirect

Int. J. Human-Computer Studies 65 (2007) 724–736

International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Automatic prediction of frustration

Ashish Kapoor[a,1], Winslow Burleson[c], Rosalind W. Picard[b,*]

[a] *Microsoft Research, 1 Microsoft Way, Redmond, WA 98052, USA*
[b] *MIT Media Laboratory, 20 Ames Street, Cambridge, MA 02139, USA*
[c] *Computer Science and Engineering/Arts, Media and Engineering, Arizona State University, 699 S. Mill Avenue, Tempe AZ, 85281, USA*

Received 15 March 2006; received in revised form 23 January 2007; accepted 13 February 2007
Communicated by Dr. S. Wiedenbeck
Available online 5 March 2007

Abstract

Predicting when a person might be frustrated can provide an intelligent system with important information about when to initiate interaction. For example, an automated Learning Companion or Intelligent Tutoring System might use this information to intervene, providing support to the learner who is likely to otherwise quit, while leaving engaged learners free to discover things without interruption. This paper presents the first automated method that assesses, using multiple channels of affect-related information, whether a learner is about to click on a button saying "I'm frustrated." The new method was tested on data gathered from 24 participants using an automated Learning Companion. Their indication of frustration was automatically predicted from the collected data with 79% accuracy (chance = 58%). The new assessment method is based on Gaussian process classification and Bayesian inference. Its performance suggests that non-verbal channels carrying affective cues can help provide important information to a system for formulating a more intelligent response.

Keywords: Affective Learning Companion; Intelligent Tutoring System; Learner state assessment; Affect recognition

1. Introduction

Expert human teachers are adept at recognizing and addressing the emotional state of learners and, based upon that observation, taking some action that positively impacts learning. For example, a savvy human tutor can discriminate whether a learner is making mistakes and does not need intervention (perhaps the learner is content to fail till he or she succeeds without aid) or whether the learner is growing increasingly frustrated with making mistakes, and is likely to quit. We would like to equip automated tutors and Learning Companions with similar perceptual abilities, so that they can be smart about when to intervene and when to leave learners to explore (and make mistakes) without being interrupted. Our hypothesis is that affective state plays a key role in discriminating these situations, and that looking at just the learner's performance on the task, e.g., characterizing mistakes, is inadequate. Thus, we examine if channels of information that carry affective cues can provide useful information to the system for identifying frustration, so that the system might ultimately formulate a more respectful response.

An Intelligent Tutoring System that responds appropriately to the negative affective states that a learner goes through while stuck on a problem can likely reengage the learner in challenging learning experiences. However, it is very hard to be sure if a learner is in a negative state or not. Learners often try to appear like they are fine when they are not (Schneider and Josephs, 1991). Self-reported feelings at the end of a task are notoriously unreliable, and getting outside observers to rate feelings is an enormous task, requiring multiple coders per section of data, agreement on meaning of labels and appearance of behavior associated with the labels, validation of the coder's ability to perceive emotion, and more. Less is known about the reliability of labeling by a person in the

[*] Corresponding author.
E-mail addresses: akapoor@microsoft.com (A. Kapoor), winslow.burleson@asu.edu (W. Burleson), picard@media.mit.edu (R.W. Picard).
[1] This work was done while the author was at the MIT Media Laboratory.

1071-5819/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2007.02.003

"heat of the moment," while they are frustrated and they have the opportunity to clearly say so. We focus this work on this case, where a learner has before her or him two different buttons labeled "I'm frustrated" or "I need some help". The learner can ignore these buttons, or click one of them. When a person clicks the "I'm frustrated button" we label the segment leading up to the click as "frustration."

Note that learners might be frustrated and might still not click the button. For example, they may feel uncomfortable confessing they feel frustrated. While we cannot know for sure how much true frustration data we miss with this button-clicking method, there are prior studies from human–computer interaction that suggest that the procedure of collecting self-perceived negative information by computer may at least be more accurate than collecting the same information from a trained expert person. For example, there is classic work whereby people are more willing to share negative information about themselves (embarrassing medical conditions, excessive drinking, etc.) with a computer than with a trained physician (Card et al., 1974; Lucas et al., 1977; Robinson and West, 1992).

Since our research agenda is particularly focused on helping learners persevere through frustration and a state of Stuck (Burleson and Picard, 2004) we use the user's self-labeling as an indication of them being frustrated and aware of it, and go back and collect their behavioral data leading up to that button click. These data are pooled against comparable data from the learners that did not indicate frustration, (i.e. those that either did not click on a button or those that clicked on "I need some help", since even though they had the opportunity to, they chose not to express frustration.) These two pools of data are used to construct an automated system that can discriminate these two classes (frustrated vs. other). The system is then tested on a set of data that was not used to train the system.

In this paper, we address the problem of recognizing the state in which a child who begins a problem-solving activity on the computer is frustrated. The scenario we focus on has a child involved in an activity (solving the Towers of Hanoi puzzle) in an environment with sensors that can measure video from the face, postural movement from the chair, skin conductance (wireless sensor on non-dominant hand), and pressure applied to the mouse. The machine combines this information in a novel way to try to infer if the child is behaving in a way that, based on prior experience of the system, has been shown to precede clicking on the frustration button.

2. Affective Learning Companions: prior art

Several researchers, including Tak-Wai Chan who coined the term "Learning Companion" with respect to Intelligent Tutoring Systems, promote the idea of presenting agents as peer companions (Chan and Baskin, 1988). One rationale for this is that peer tutors can be effective role models because they are less likely to invoke anxiety in learners; learners may believe they can attain the same level of expertise as their tutors, while they may not believe that they can attain an adult teacher's level of expertise. Regulation of anxiety and other negative feelings is increasingly recognized as important in successful learning experiences. Researchers Schank and Neaman (2001) acknowledge that fear of failure is a significant barrier to learning and believe this can be addressed in several ways: minimizing discouragement by lessening humiliation; developing the understanding that consequences of failure will be minimal; and providing motivation that outweighs or distracts the unpleasant aspects of failure. Because failure is important to learning and instrumental to the development of multiple points of view required for deep understanding, we think it is critical that learning systems not only let learners fail, but encourage them to persevere in the face of failure. However, it is also critical that they help learners manage the negative feelings associated with failure. Perseverance through failure can be turned into learning; it does not have to lead to the intense frustration that often results in quitting or in a future desire to avoid similar experiences.

Through the integration of advanced graphics and sensing environments with rich content and increasingly sophisticated behavior repertoires, research efforts are increasing participants' sense of presence and social engagement with agents in immersive environments and in learning environments (Lester et al., 1999; Johnson et al., 2003). In his work on the development of animated agents for learning, Lester demonstrated a phenomena he has termed the persona effect, which is that "the presence of a lifelike character in an interactive learning environment—even one that is not expressive—can have a strong positive effect on student's perception of their learning experience" (Lester et al., 1997). Through the use of verbal and non-verbal affective communication, Intelligent Tutoring System and Learning Companion technologies are moving beyond the capabilities that gave rise to the persona effect, toward realizing real-time multi-modal affective interactions between agents and learners. However, to date, there are no examples of agents that can fully sense natural (both verbal and non-verbal) human communication of emotion and respond in a way that rivals that of another person.

Existing agent systems typically infer human affect by sensing and reasoning about the state of a game or an outcome related to an action taken by the user within the computing environment. Use of such an approach is illustrated by the pedagogical agent COSMO, who applauds enthusiastically and exclaims "Fabulous!" if the student takes an action that the agent infers as deserving of congratulations (Lester et al., 1999). There are learning situations in which this reaction would be warmly received and perhaps reciprocated with a smile by the user, and situations such as when a student is bored, frustrated and ready to quit where it most certainly would not. While reasoning based on a user's direct input behaviors is important and useful, it is also limited. For example, COSMO has no ability to see how the user responds non-verbally to its enthusiasm. COSMO is unable to tell, for example, if the user beamed with pride or frowned and rolled her eyes, as if to say that COSMO's response was excessive or otherwise inappropriate. If the latter, it might be valuable for COSMO to acknowledge its gaffe, thus making it less likely the user will hate it or ignore it in the future. In addition to providing better social interactions, understanding the learner's affect might enable the Intelligent Tutoring System to help learners address and better understand the role of their own feelings in the learning process. This is discussed more in Section 2.1. Thus, there is a need and a desire to advance agent capabilities to include perceptual sensing of non-verbal affective expressions together with the channels that are traditionally sensed in interactive agent systems.

Affect can be expressed in many ways—not just through voice, facial expressions and gestures, but also through the adverbs of many aspects of the interaction. Affect modulates how a learner types and clicks, what words are chosen and how often they are spoken. It also impacts how a learner fidgets in the chair and moves head and facial muscles. In the development of an affective Intelligent Tutoring System, a promising approach is to integrate many channels of information in order to better understand how affect is communicated. Physiological and affective sensors (including sensors that measure heart rate, skin conductance, elements of respiration, blood oxygen levels, pressure exerted on a mouse, posture in a chair, gait analysis, brain oxygen levels, etc.) are emerging as new technologies for human agent interactions (Picard, 1997, 2000; Allanson and Fairclough, 2004). There are many challenges to the design and use of multi-modal affective sensors. For ease of use and natural interactions, it is desirable to have systems that are not intrusive and do not require any training. Individual sensors also have signal-to-noise issues and robustness issues. Then there are the reactions of the users to the sensors, in terms of ethical issues, privacy, and comfort (Reynolds, 2005). Some researchers have found that children do not find a skin conductance sensor to be intrusive; in fact if kids are "deprived" of the opportunity to use the sensor, they reportedly felt they had missed out on part of the experience of the interaction (Conati, 2004).

An emerging approach to Intelligent Tutoring System development is to assess and attend to learners' affective states with the assistance of sensors. Pattern recognition with multi-modal sensors has been shown to be an effective strategy in the development of affective sensing. Research incorporating the posture analysis seat (Mota and Picard, 2003) and the Blue-Eyes camera (Haro et al., 2000) has classified engagement, boredom and break-taking behavior with 86% accuracy (Kapoor and Picard, 2005). Another project that takes this approach is the AutoTutor project at the University of Memphis (D'Mello et al., 2005). This project incorporates a posture chair, facial expression camera, and conversational cue analysis to inform agent interactions with college students. Once a correlation is made between learners' affective states and the sensor values, this correlation can inform agent's interactions enabling them to become responsive to learners' affect in real time. The work in this paper advances these prior works by employing additional sensors for greater reliability, using a new technique for combining the information, and by using a strategy that obtains users' self-labeling of their state of frustration.

Finally, we would also like to point out that there has been some work on detection of frustration in scenarios other than Learning Companions. Fernandez and Picard (1998) demonstrated signal processing techniques for recognizing frustration in users from galvanic skin response (GSR) and blood volume pressure (BVP). Qi et al. (2001) focused on detecting frustration in users filling out web forms by using only a pressure sensitive mouse. Similarly, Mentis and Gay (2002) and McLaughlin et al. (2004) have used haptic sensors to detect frustration.

2.1. *The affective intervention*

Results from studies of human–human interaction can usefully inform the design of human–computer interaction, e.g. Reeves and Nass (1996) and Moon (2000). We thus turn to the literature on human interaction in order to develop hypotheses about what is likely to succeed or fail in human–computer interaction. For example, the research of Robinson and Smith-Lovin (1999) illustrates the importance of an appropriate affective response: their work describes how if a person responds positively to something bad happening, then that person will be less liked. Alternatively, if a person responds in a way that is affectively congruent, then that person will be more liked. These findings seem to support the current approach in pedagogical agent research where the character smiles when you succeed and looks disappointed if you make a mistake or fail. However, without additional information from sensor channels that classify elements of affect, these systems run the risk of making mistakes that may negatively impact the learner and their feelings about the system. Human tutors do not reliably smile every time you do something right, or frown every time you make a mistake, but they do often display an expression that is empathetic with yours.

Lepper et al. (1993) have found that approximately 50% of expert tutors' interactions with their students are affective in nature. For example, at times of frustration, or perhaps when a learner appears ready to quit, expert tutors might be empathetic and encourage a student using a strategy such as Dweck's (the mind is like a muscle—it can get stronger as you work it). At other times, when the likelihood of quitting is low, the tutor or system's interactions might choose not to interrupt the learner's experience, letting the learner self-regulate.

In the field of Intelligent Tutoring Systems there is a distinction between, on the one hand, adjusting the environment or task to facilitate uninterrupted flow

(Malone, 1984; Hill et al., 2001), and on the other, empowering the user through self-awareness to participate in self-regulated motivational strategies. Many Intelligent Tutoring Systems choose to adjust the challenge level to keep the learner engaged. If a learner is frustrated, these systems attempt to make the activity easier. This approach does not address a learner's emotional state. In contrast, through social interactions, we choose to help individuals tailor self-perceptions of their ability with respect to a challenge. If learners are able to alter their perception of failure and negative affect, then they may be able to mitigate the detriments of negative asymmetry. This may enable them to persevere and succeed at greater challenges. With an increased awareness of and ability to manage their negative affect, students are likely to reengage in challenging learning experiences in the future.

Carol Dweck's work on self-theories of intelligence presents promising findings for understanding why learners fail and how to help them succeed. She has found that individuals' beliefs of their own intelligence profoundly affect their motivation, learning, and behavioral strategies, especially in response to their perception of failure (Dweck, 1999). This research has identified two predominant groups of individuals: "incrementalists," who believe their own intelligence can be enhanced, and "trait learners," who believe their intelligence is largely fixed. She has found that when incrementalists fail at a task, they tend to increase their intrinsic motivation for the task, believing that if they try harder, they will get better and smarter. When trait-based individuals fail, they exhibit avoidance and decreased intrinsic motivation for the task, believing instead that their previous performance defines their ability. They act on their desires to avoid further confirmation of what they perceive to be their "trait-based" inability; they tend to quit at the first signs of difficulty. She has developed a simple strategy of metacognitive knowledge, a strategy for thinking about thinking: the strategy is to recall that "the mind is like a muscle and through exercise and effort you can grow your intelligence." By adopting this simple strategy people can shift their self-theories of intelligence.

Dweck's work on self-theories of intelligence uses interventions at a group level. Our research opens new possibilities by setting the stage for sensing and responding to learners on a personal level, identifying where each learner is getting frustrated, and then trying and evaluating various interventions timed specifically for that individual's needs.

2.2. Interaction strategies once frustration is detected

This paper is focused on predicting when a learner is frustrated, by examining data leading up to the time when he or she clicks on the "I'm frustrated" button. In work that follows this, we intend to develop individually tailored interventions based on Dweck's strategy. We seek to help learners understand and use their frustration as an indication of a learning opportunity. It is important to note that one of the advantages of incorporating affective interactions into Intelligent Tutoring Systems is that tailoring the agent's responses to individual learner's interactions can be done in a reliable and controllable manner. By reliable we mean these interactions can be coordinated by detection algorithms that yield consistent results that are not subject to the typical forms of experimenter bias in human–human interactions. By controllable we mean that all the elements of an agent's expression type, intensity, duration, etc. can be computationally adjusted, repeated, recorded, and analyzed across all experimental participants. For example, if we want to test if mirroring a particular behavior influences some outcome, we can test it more easily with an automated mirroring system than by asking a human tutor to mirror (since people do not usually have reliable control over their non-verbal communication).

3. The ALC architecture

We developed a novel platform, shown in Fig. 1, for affective agent research. It integrates an array of affective sensors in a modular architecture that drives a system server and data logger, inference engine, behavior engine, and character engine. The character engine includes dynamically scripted character attributes at multiple levels. This approach is particularly suited to communication of affective expression. The user sits in front of a wide screen plasma display. On the display appears an agent and 3D environment. The user can interact with the agent and can attend to and manipulate objects and tasks in the environment. The chair that the user sits in is instrumented with a high-density pressure sensor array. The mouse detects applied pressure throughout its usage. The user also wears a wireless skin conductance sensor on a wristband with two adhesive electrode patches on the hand. Two cameras are in the system, a video camera for offline coding and the Blue-Eyes camera to record elements of facial expressions. This multi-modal approach to recognizing affect uses more than one channel to sense a broad spectrum of information. This approach applies techniques from psychophysiology, emotion communication, signal processing, pattern recognition and machine learning, to make an inference from this data. Since any given sensor will have various problems with noise and reliability, and will contain only limited information about affect, the use of multiple sensors should also improve robustness and accuracy of inference.

In addition to non-verbal interactions, the character interacts with the user through an asynchronous voice dialogue (Burleson et al., 2004). The character speaks using Microsoft's "Eddie" voice scripted with Text-Aloud, a text-to-speech application. When there are questions the words are presented in a text bubble, as well, for the user to read. Users may respond by clicking on the available text responses. Finally, the video camera records the user and the onscreen activity. It is positioned to acquire both an

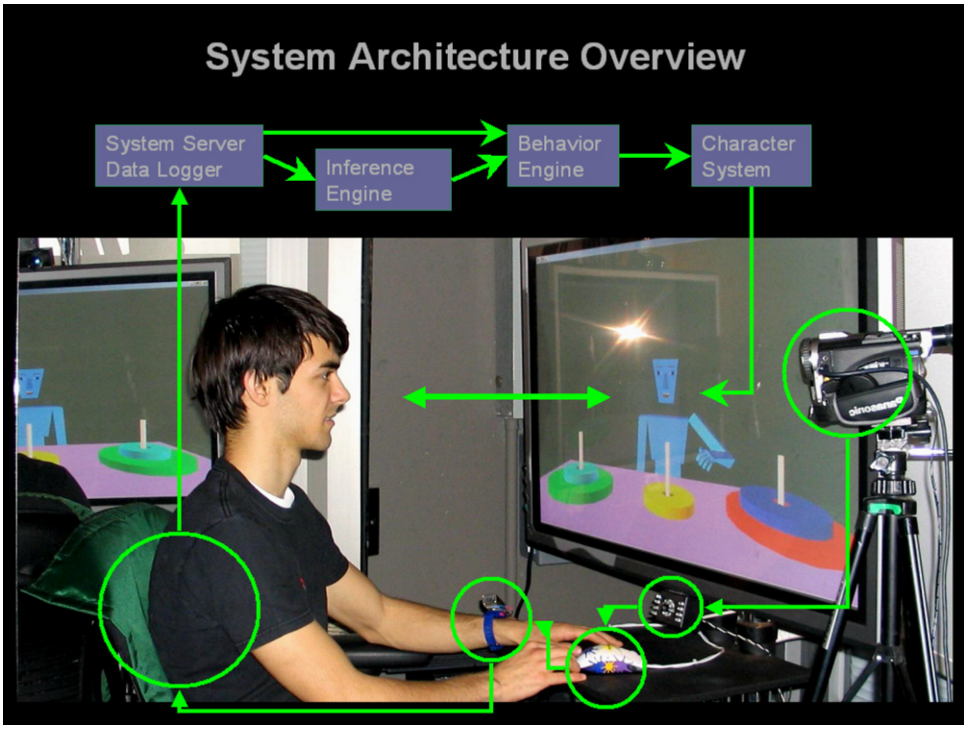


Fig. 1. System architecture with (from right to left): video camera, Blue-Eyes camera, pressure-sensitive mouse, skin conductance sensor, and pressure-sensitive chair.

image of the user and an image of the screen that is reflected in a mirror positioned behind the user's head. This arrangement was chosen so as not to miss any of the features of the user/character interaction while providing true (same image) synchronization. When the system is initialized, a datagram signal is sent to start the DirectX video capture and the time is noted in the log. The rest of this section describes the sensing, the pattern recognition to detect the affective markers of the "frustration" state and the challenges in designing appropriate affective responses for the character.

3.1. The sensing

The non-verbal behaviors are sensed through a camera, a pressure sensing chair, a pressure mouse and a device that measure the skin conductance. Here we describe them briefly.

The Blue-Eyes camera: We use an in-house built version of the IBM Blue-Eyes Camera that tracks pupils unobtrusively using structured lighting (Haro et al., 2000). The system exploits the red-eye effect to track pupils. Once tracked, the pupil positions are passed to a method to detect head nods and shakes based on hidden Markov models (HMMs) (Kapoor and Picard, 2001). This method provides the likelihoods of nods and shakes. Similarly, we have also trained an HMM that uses the radii of the visible pupil as inputs to produce the likelihoods of blinks. Further, we use another system we developed earlier to recover shape information of eyes and eyebrows (Kapoor and Picard, 2002).

Given pupil positions we can also localize the image around the mouth. Rather than extracting the shape of the mouth explicitly we extract two real numbers that correspond to two kinds of mouth activities: smiles and fidgets. We look at the sum of the absolute difference of pixels of the extracted mouth image in the current frame with the mouth images in the last 10 frames. A large difference in images is treated as mouth fidgets. Besides a numerical score that corresponds to fidgets, the system also uses a support vector machine (SVM) to compute the probability of smiles. Specifically, an SVM was trained using natural examples of mouth images, to classify mouth images as smiling or not smiling. The localized mouth image in the current frame is used as an input to this SVM classifier and the resulting output is passed through a sigmoid to compute the probability of smile in the current frame. The system can extract features in real time at 27–29 frames per second on a 1.8 GHz Pentium 4 machine. The system tracks well as long as the subject is in the reasonable range of the camera. The system can detect whenever it is unable to find eyes in its field of view, which might occasionally happen due to large head and body movements. We found that many children do move a lot, so it is important to have a system that is robust to natural movement.

The posture sensing chair: Postures are recognized using two matrices of pressure sensors made by Tekscan. One matrix is positioned on the seat-pan of a chair; the other is placed on the backrest. Each matrix is 0.10 mm thick and consists of a 42-by-48 array of sensing pressure units distributed over an area of 41 × 47 cm. A pressure unit is a variable resistor, and the normal force applied to its superficial area determines its resistance. This resistance is transformed to an 8-bit pressure reading, which can be interpreted as an 8-bit grayscale value and visualized as a grayscale image. Fig. 2(b) shows the feature extraction strategy used for postures in Mota and Picard (2003). First, the pressure maps sensed by the chair are pre-processed to remove noise and the structure of the map is modeled with a mixture of Gaussians. The parameters of the Gaussian mixture (means and variances) are used to feed a three-layer feed-forward neural network that classifies the static set of postures (for example, sitting upright, leaning back, etc.) and activity level (low, medium and high) in real time at 8 frames per second, which are then used as posture features by the multi-modal affect classification module.

Pressure mouse: The pressure mouse has eight force-sensitive-resistors that capture the amount of pressure that is put on the mouse throughout the activity (Reynolds, 2001). Users who found an online task frustrating have been shown to apply significantly more pressure to a mouse than those who did not find the same task frustrating (Dennerlein et al., 2003).

Wireless BlueTooth skin conductance: A group at Media Lab Europe and at the MIT Media Lab, Strauss et al. (2005) developed a wireless version of an earlier "glove" that senses skin conductance. While the skin conductance signal does not explain anything about valence—how positive or negative the affective state is—it does tend to be correlated with arousal, or how activated the person is. High levels of arousal tend to accompany significant, new, or attention-getting events.

Game state: While game state is not a traditional sensor, and is not used in this paper, it is gathered by our system as a source of data and is treated as a sensor channel in a manner similar to each of the other sensors. Thus, if desired, it can be included in any decision-making about learner state and appropriate interventions. The system records the disk state after each move, checks if it is legal or illegal, increments the move count, calculates the optimal number of moves to the end of the game (Rueda, 1997), and evaluates progress in terms of number and significance of regressions. This data can also potentially be used to explore users' engagement and intent: understanding of the game, proceeding in a focused way, or becoming disengaged.

3.2. *Decision making: detecting frustration*

The data observed through the sensors are classified into "pre-frustration" or "not pre-frustration" behavior based on probabilistic machine learning techniques described in the rest of this section.

Gaussian process classification: Gaussian process (GP) classification (Williams and Barber, 1998) is a technique for pattern recognition and machine learning that has been shown to outperform popular techniques such as SVMs in many cases. Let $\mathbf{X} = \{\mathbf{X}_L, \mathbf{X}_U\}$ consist of the data that have been labeled, $\mathbf{X_L} = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\}$ and the data that are still unlabeled $\mathbf{X}_U = \{\mathbf{x}_{n+1} \ldots \mathbf{x}_{n+m}\}$. The given (known) labels are denoted by $\mathbf{t}_L = \{t_1, \ldots, t_n\}$ and the random variables for the class labels still to be determined by the algorithm are denoted by $\mathbf{t}_U = \{t_{n+1}, \ldots, t_{n+m}\}$. We are thus interested in obtaining the distribution of the labels we do not know, given everything else that we have observed: $p(\mathbf{t}_U|\mathbf{X}, \mathbf{t}_L)$. In this paper we limit ourselves to two-way classification, where the labels correspond to either pre-frustration (data leading up to clicking on "I'm frustrated") or not pre-frustration (a similar window of data but for those people who clicked on "I need some help" or who never clicked on a button).

Intuitively, the idea behind GP classification is that the hard labels $\mathbf{t} = \{\mathbf{t}_L, \mathbf{t}_U\}$ depend upon hidden soft-labels $\mathbf{y} = \{y_1, \ldots, y_{n+m}\}$. These hidden soft-labels arise from a

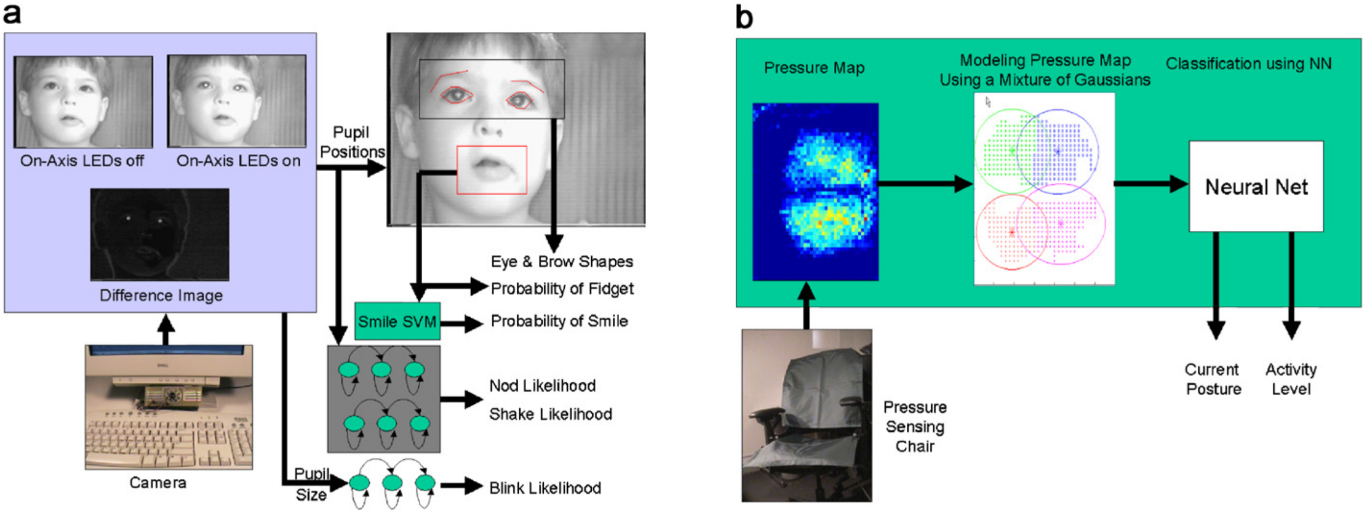


Fig. 2. Modules to extract (a) facial features and (b) posture features.

GP which in turn imposes a smoothness constraint on the possible labelings. Given the labeled and unlabeled data points, our task is then to infer $p(\mathbf{t}_U|D)$, where $D = \{\mathbf{X}, \mathbf{t}_L\}$. Specifically:

$$p(\mathbf{t}_U|D) = p(\mathbf{t}_U|\mathbf{X}, \mathbf{t}_L) \propto \int_{\mathbf{y}} p(\mathbf{t}_U|\mathbf{y})p(\mathbf{y}|\mathbf{X}, \mathbf{t}_L). \qquad (1)$$

The full Bayesian treatment of GP classification requires computing the integral given in Eq. (1). The key quantity to compute is the posterior $p(\mathbf{y}|\mathbf{X}, \mathbf{t}_L)$, which if obtained in a simple approximation (such as Gaussian) can be used to perform the Bayesian averaging as in Eq. (1). Alternatively, we can use the mean (Bayes point) of the posterior to classify any unseen point. The details on the Bayes point classification can be found in Herbrich et al. (2001) and in Minka (2001b). Note, the required posterior can be written as

$$p(\mathbf{y}|\mathbf{X}, \mathbf{t}_L) = p(\mathbf{y}|D) \propto p(\mathbf{y}|\mathbf{X})p(\mathbf{t}_L|\mathbf{y}). \qquad (2)$$

The term $p(\mathbf{y}|\mathbf{X})$ in Eq. (2) imposes a smoothness constraint such that the solutions that have the same labelings for similar data points are preferred. The similarity between the data points is defined using a function called a kernel. Examples of kernels include Gaussian functions, polynomial functions, etc. In this work we define the similarity using a weighted Gaussian kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left[-\frac{1}{2} \cdot \sum_{k=1}^{K} \frac{(x_i^k - x_j^k)^2}{\sigma_k^2}\right]. \qquad (3)$$

Here, x_i^k corresponds to the kth feature in the ith sample. Note that we are weighing each dimension individually before exponentiating. By adjusting the value of the parameter σ_k, we can control the contribution of the kth feature in the classification scheme. A priori we do not have any information regarding which features are most discriminatory. The classification performance highly depends upon finding the right set of feature weighting parameters, and this issue is addressed in the next section.

A smoothness constraint is imposed by assuming a GP prior over the soft labels $\mathbf{y} = \{y_1, \ldots, y_{n+m}\}$. Consequently, the soft labels $\mathbf{y}$ are assumed to be jointly Gaussian and the covariance between two outputs y_i and y_j is specified by the kernel function applied to $\mathbf{x}_i$ and $\mathbf{x}_j$. Formally, $\mathbf{y} \sim \mathcal{N}(0, K)$ where K is a $(n+m)$-by-$(n+m)$ kernel matrix with $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. The entries in the matrix K capture the notion of similarity between two points. Note that the formulation of this prior can be extended to any finite collection of data points.

The second term $p(\mathbf{t}_L|\mathbf{y})$ in Eq. (2) is called the likelihood, and it incorporates information provided in the labels. Let $\Phi(\cdot)$ be a step function. The observed labels $\mathbf{t}_L$ are assumed to be conditionally independent given the soft labels $\mathbf{y}$ and each t_i depends upon y_i through the conditional distribution:

$$p(t_i|y_i) = \Phi(y_i \cdot t_i) = \begin{bmatrix} 1 & \text{if } y_i \cdot t_i \geqslant 0 \\ 0 & \text{if } y_t \cdot t_i < 0 \end{bmatrix}.$$

Due to the form of this likelihood, computing the posterior $p(\mathbf{y}|\mathbf{X}, \mathbf{t}_L)$ is non-trivial. In this work we use expectation propagation (EP), a technique developed by Minka (2001a) for Bayesian inference, to approximate the posterior $P(\mathbf{y}|D)$ as a Gaussian distribution. Conceptually, we can think of EP starting with the Gaussian prior $\mathcal{N}(0, K)$ over the hidden soft labels $(\mathbf{y})$ and then approximating the non-Gaussian likelihood terms as a Gaussian distribution using a message passing scheme. For details readers are encouraged to look at Minka (2001a).

Learning the similarity function: The performance of the algorithm depends highly upon the kernel widths, $[\sigma_1, \ldots, \sigma_K]$, because they define the similarity measure between two data points. Finding the right set of all these parameters can be a challenge. Many discriminative models, including SVMs often use cross-validation, which is a robust measure but can be prohibitively expensive for real-world problems and problematic when we have few labeled data points.

In our work, we choose parameters that maximize the marginal likelihood or the evidence, which is the same as the constant $p(\mathbf{t}_L|\mathbf{X})$ that normalizes the posterior. This methodology of tuning the parameters is often called evidence maximization and has been one of the favorite tools for learning the parameters of the kernel function. Evidence is a numerical quantity and signifies how well a model fits the given data. By comparing the evidence corresponding to the different models, we can choose the model with the parameters suitable for the task. (Note: in machine learning, these parameters that are learned for different models are often called *hyperparameters*.)

Let us denote the set of kernel parameters as $\Theta = \{\sigma_1, \ldots, \sigma_K\}$. Formally, the idea behind evidence maximization is to choose a set of parameters that maximize the evidence. That is, $\hat{\Theta} = \arg\max_{\Theta} \log [p(\mathbf{t}_L|\mathbf{X}, \Theta)]$. When the parameter space is small then a simple line search or the Matlab function **`fminbnd`**, based on golden section search and parabolic interpolation, can be used. However, in the case of a large parameter space exhaustive search is not feasible. In those cases, non-linear optimization techniques, such as gradient descent or expectation maximization (EM) can be used to optimize the evidence. In this work, we use the EM–EP algorithm (Kim and Ghahramani, 2004) to maximize evidence. Each round of EM–EP take $O(n^3)$ steps, where n is the number of data points in the data set. For the classification problem on 24 subjects, we have 24 points in the data set and the whole algorithm takes around 2–3 min to converge during training. Once we have finished training the algorithm, i.e., we have the kernel parameters and the Bayes point, then classifying a new sample of data takes only a few milliseconds.

4. Experimental evaluation

We performed experiments on a set of natural data from 24 middle school students aged 12–13 to evaluate how well the proposed system worked. The task was a six-disk version of the Towers of Hanoi puzzle. The details of data collection and rest of the experiment are discussed below.

4.1. Data collection

The methodology followed the timeline presented in Table 1.

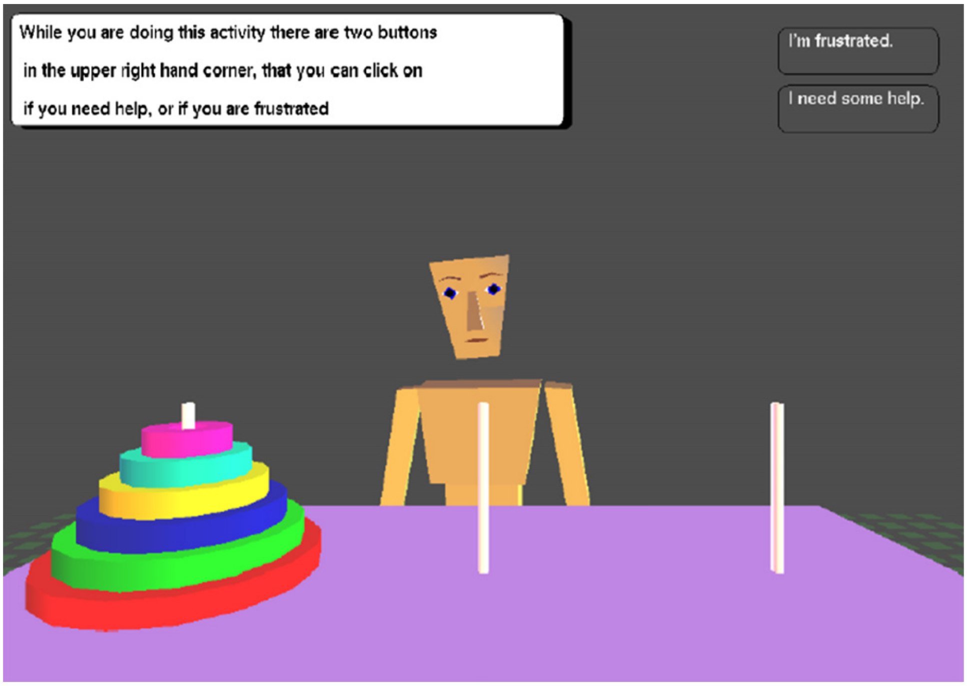
First, a pre-test was administered to determine the learner's self-theory of intelligence and their goal mastery orientation (Dweck, 1999). Next, the participants were shown a 7 min slide show based on a script that Dweck has used to beneficially shift children's beliefs about their own intelligence. Next, the agent appeared and presented the Towers of Hanoi activity. Throughout the time of the activity, there were two buttons prominent at the top of the screen that users could click on: "I'm frustrated" and "I need some help" (see Fig. 3). The user was free to ignore these buttons or to click one at any time.

If a learner clicked on one of the buttons, or after 16 min, whichever occurred first, the learner was presented with a supportive dialogue by the character during which he or she is encouraged to continue. After 16 min (from the start of the activity) or when he or she finished the activity a post-activity survey was administered with questions about the experience followed by the modified working alliance inventory used to gauge learners' impressions of the character (earlier work has used this instrument to assess bonding between a person and an agent Bickmore, 2003). A second activity was then administered followed by the post-test self-theory of intelligence and goal mastery orientation surveys, and a debriefing.

From the participant's perspective they are asked to fill out several pages of the pretest survey, then to watch a 7 min slide show that discusses strategies for learning and becoming more intelligent. The slide show was a full screen Power Point presentation narrated by a recorded human voice. These slides were presented to participants on the same monitor as they would subsequently see the agent. Then the agent appears standing behind three poles with six disks on the left-most pole. The agent introduces itself, the activity, and asks if they have seen this before. It also presents the two buttons and encourages them to start the activity, saying, "While you are doing this activity there are two buttons in the upper right hand corner, that you can click on if you need help, or if you are frustrated. Click on a disk to start, whenever you want. I'll just watch and help if I can." At first the activity may seem to be simple and perhaps even fun. Quickly it becomes apparent that it is much more challenging than it looks and it can feel quite difficult to make progress. The task involves recursion, which can make you feel like you are going backwards when in fact you are going forwards. After several minutes of repeated attempts to make progress, many participants lose their motivation and are ready to end this experience. As they get to this stage they may click on one of the buttons. When they click on one of the two buttons the character reminds them of a message from the slide show saying, "I'm sorry I don't know more about this activity so I could help you through it, but remember the mind is like a muscle, and just like when you exercise your muscles and they get stronger, you can exercise your mind and increase your intelligence. If you stick with it you can get better and stronger by learning". After one of the two buttons has been clicked, the buttons are no longer displayed, and participants are given up to 16 min from the time they started the activity, to finish the activity. They are presented with the post-activity survey and modified working alliance inventory. If they do not click on one of

Table 1
Experiment protocol with durations in minutes; the approximate values indicate that these events have participant interactions and therefore some variation in duration

Protocol events for all subjects	Duration in minutes
Assent and consent forms	~3
Initial survey questions and pre-test (including self-theories of intelligence and goal mastery orientation)	~10
Slide presentation (based on Dweck's message)	~7
Agent appears and introduces activity and the two buttons	~2
Participant engages in Towers of Hanoi activity with option to press either button	~16
If participant presses either button character provides supportive dialogue	~1 during the 16 min
Participant continues to engage in Towers of Hanoi task	Until completing the activity or until 16 min from the start of the activity
Post-activity survey of experience	~1
Modified working alliance inventory	~3
Participants are presented with second activity (Rush Hour Puzzle)	Until completing the activity or until 10 min from the start of the second activity
Post-test (including self-theories of intelligence and goal mastery orientation)	~10
Debrief	~2



While you are doing this activity there are two buttons in the upper right hand corner, that you can click on if you need help, or if you are frustrated

Fig. 3. Learner views this screen after the character explains it is time to start the Towers of Hanoi task. Either of the two buttons at upper right can be pushed at any time.

Table 2
Fourteen features used for classification of learner state

Face tracker	Posture sensor	Skin conductance	Pressure mouse
Presence of nods	Activity	Conductance	Activity
Presence of shakes	Ratio of postures		Skew
Probability of fidget			Mean
Probability of smile			Variance
Presence of blink			
Head velocity			
Head tilt			

the buttons and instead persevere in the activity for 16 min they are presented with the post-activity survey and modified working alliance inventory at that time. After a second activity and the post-test self-theories of intelligence and goal mastery orientation surveys, they are given a debriefing with the opportunity to ask questions.

4.2. Data preprocessing

We collected sensory data for each subject during the course of interaction with the character. The raw data from the camera, the posture sensor, the skin conductance sensor and the pressure mouse is first analyzed to extract 14 features that are summarized in Table 2. The pressure mouse data were processed to obtain the following features: activity, skew, mean, and variance (Reynolds, 2001). This set of data did not contain any unusual noise and so those features were extracted directly. The Wireless BlueTooth skin conductance data exhibited occasional spikes and sudden drops in values and these were filtered to eliminate any values that exceeded 5 standard deviations from the average. The posture analysis seat data did not exhibit any significant noise in obtaining the following features: ratio of forward to backward posture, activity level. The Blue-Eyes camera system data were filtered so that only data that detected both pupils were included. The following features were obtained: presence of head nod, presence of head shake, probability of fidget, probability of smile, presence of blink, head velocity, and head tilt.

For the purpose of classification we use the values averaged over 150 s of activity. We tried various window sizes and the window size of 150 s gave the best results. All the features are normalized to lie between the values of zero and one before we take the mean. For the children that clicked on the frustration button, we consider samples that summarize 150 s of activity preceding the exact time when they clicked. However, for the children that did not indicate they were frustrated, we needed to come up with a comparable 150 s window. We examined the timing of the frustration clicks and observed that most subjects that clicked on the frustration button did so before 375 s, with the average time of clicking at around 174 s. Further, to analyze the non-frustration group we first look at the clicking time for the cases where the frustration button was pressed after 100 s, allowing some significant interaction before ending the session. The median among these longer clicking times was 225 s. Considering this, and the duration of the data in the non-frustration group, we decided to use a 150-s window beginning 225 s after the start of the game for the non-frustration group. Out of 24 children, 10 clicked the frustration button; thus, we have a data set with 24 samples of dimensionality 14 (because of the 14 features), where 10 samples belong to class +1 (frustrated) and 14 to class −1 (not frustrated).

4.3. Classification results

We performed experiments using different classification techniques, which include the classic techniques of one-nearest neighbor and SVMs, as well as the newer technique of GP classification. Table 3 shows the classification results. The accuracies are reported using the leave-one-subject out strategy, which is to first choose all but one subject's labels as training examples and test the performance on the subject who was left out. This process is repeated for all the available samples and we report all the correct detections and misses as well as the false alarms (Table 3).

With GP classification we use Eq. (3) as a kernel, which weighs all the features individually and we exploit the evidence maximization framework to tune those weights. Thus, GP classification allows us to use expressive models that can capture various relationships between the features in the data. This is valuable since we do not know up front which features are going to be most discriminatory. The dimensionality of the observations is 14 in this example and it is non-trivial to tune these weights using cross-validation. For SVMs, we stick with the (commonly used) radial basis function (RBF) kernel and use leave-one-out cross-validation to tune the kernel width. Table 3 demonstrates the advantage of using the Bayesian framework and we can see that the GP classification outperforms both one-nearest neighbor strategy and the SVM with an RBF kernel. We also compare to a control method that represents random chance. Suppose we had an equal number of subjects who did and did not click on the frustration button; then, we could set chance to 50%. However, since we know for this data that $\frac{14}{24} = 0.583$ of the subjects did not click on the frustration button, we could simply assign everyone to this category and be right 58.3% of the time. We thus use this higher number to represent the chance condition.

Further, we also train an SVM that uses the maximum evidence kernel learnt using EM–EP algorithm for the GP classification and as we can see it performs the same as GP classification. Note that in this case the EM–EP algorithm has already found the correct kernel for classification and SVM is greatly benefiting from this computation. Here, the SVM does not compute the optimal kernel but only uses the kernel parameters learnt via the EM–EP algorithm that help the GP to discriminate the data well. Nonetheless, the results are the same (79.17% for both the methods) once you have computed this kernel. This experiment illustrates that GP classification provides not only a good framework to classify these data, but also an additional benefit of being able to learn the parameters weighing the features through evidence maximization.

Finally, we can also look at the optimized parameters $[\sigma_1, \ldots, \sigma_K]$ to determine the most discriminative features. Fig. 4 shows the MATLAB boxplot of the kernel parameters corresponding to the different features obtained during the 24 leave-one-out runs of the EM–EP algorithm. A low value of the kernel parameter σ corresponds to *high* discriminative capability of the feature.

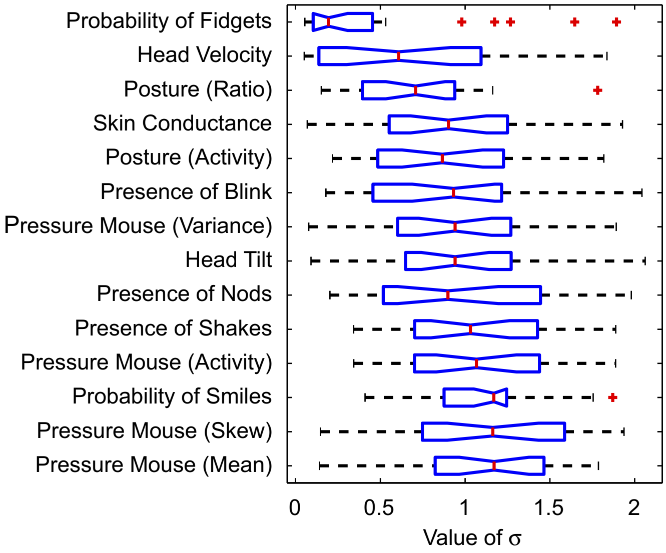


Fig. 4. Finding the most discriminative features. The MATLAB boxplot of the kernel parameters (σ) optimized during the 24 leave-one-out runs. The lines in the middle of the box represent the median, the bounding box represents quartile values, the whiskers show the extent of the values and the '+' represent the statistical outliers. A low value corresponds to high discriminative power of the feature.

Table 3
Classification results

	Clicked frustration button 10 Samples		Persevered on task or clicked help button 14 Samples		Accuracy (%)
	Correct	Misses	Correct	Misses	
Random (Control)	0	10	14	0	58.3
1-Nearest neighbor	6	4	10	4	66.67
SVM (RBF Kernel)	6	4	11	3	70.83
Gaussian process	8	2	11	3	79.17
SVM + kernel of Gaussian process	8	2	11	3	79.17

From the boxplot we can see that the fidgets, velocity of the head and the ratio of the postures are the three most discriminative features. However, there are a lot of outliers for the fidgets which implies that the fidgets can be unreliable possibly due to sensor failure and individual differences. Note, that all the parameters span a range of values suggesting that despite some statistical trends the discriminative power of a feature might depend on each individual learner. Thus, it might be beneficial for the Learning Companion to discover the patterns in user behavior rather than follow some pre-programmed rules.

5. Discussion and conclusions

In this paper we examined whether there was information in non-verbal multi-modal data that could predict if a learner was going to click on a button to say "I'm frustrated." A second button that users could press said "I need some help". We inherently assume in this case that the participants were not strongly frustrated (see below and Table 4 for details). These buttons were worded in a non-threatening way, so that people would be reasonably comfortable using them; thus we do not expect that the state of frustration that people were in before clicking on the frustration button would include strong negative outbursts (as if they had been offended or angered), but rather that they would look fairly subtle. Our goal was to see if it might be possible to classify behaviors leading up to user's clicking on the frustration button, so that future agents might consider adapting their intervention techniques based on detecting such behaviors. The experiment was the first of its kind, using a variety of custom sensors and algorithms comprising the first such system that can sense and respond in real time to a learner's multi-modal non-verbal expressions that precede frustration.

Of the 24 subjects from whom we gathered reliable multi-modal data, nine persevered without clicking a button, five clicked the help button, and 10 clicked the frustration button. Of these 24, 11 were boys and 13 were girls, with four boys and six girls clicking on "I'm frustrated" and two boys and three girls clicking on "I need some help." Neither in the participants that were frustrated, nor in the other five subjects that clicked on "I need some help", was there an obvious relationship between gender and their classification likelihood of being frustrated. Further, one of the assumptions we made in this work was that the users that clicked on the "I need some help" button were probably not strongly frustrated. Table 4 examines this assumption more carefully. This table shows the probability that a subject's 150 s window was recognized as frustration. We see that one female received a 0.77, so she would have been classified as frustrated (even though our labeled data considered her not frustrated) and one male received a 0.54, which can be considered as a borderline case.

Table 4
For each subject that clicked on the "I need some help" button, here is the algorithm's assessment of the probability he or she was frustrated

Probability that frustrated	Gender
0.20	Male
0.77	Female
0.54	Male
0.17	Female
0.22	Female

It should be noted that generalization of this work is likely to be affected by the age of the users, availability and robustness of the various sensors (e.g. the camera sensor is effected by lighting and the skin conductance sensor is effected by sweat), and even the dialogue and situation the users are presented with is likely to affect the results of classification. Thus at this stage of research in the multi-modal classification of learner's frustration, training of the classifier with data from the specific context users are presented with is likely to be necessary.

One weakness of our method is that some of the subjects clicked the buttons so soon that there was not enough time to get multiple samples of data from them. Thus, we got one window of data before they clicked on a button, and nothing more. Hence, all of our examples of "frustrated" come from a different group of people than the examples of "not frustrated" behavior. Because of this limitation there is the possibility that our algorithm is discriminating not the state that preceded clicking vs. not-clicking, but rather some other aspect that differs between the two populations. Further work is needed to be certain that our method generalizes to more than the 24 subjects here.

We applied a relatively new machine learning technique, GP classification, to the problem of learning different models, and found that it outperformed two more classical popular techniques. We thus demonstrated assessment of what appears to be a kind of "pre-frustration" state using only non-verbal cues of postural movements, mouse pressure, skin conductance, facial movements and head gestures. While these signals alone did not provide perfect classification of behavior, they did significantly outperform a random classifier (79% vs. 58%). We also do not know how accurately human observers would perform given the same task, which would be an interesting future investigation to see if they beat our algorithm or vice versa. Our result suggests that there is valuable information in these non-verbal channels that can be useful for agents and other affective systems to ascertain in deciding when to intervene.

Acknowledgments

We would like to thank Selene Mota for her software for the chair posture pattern analysis, and also many past members of the Affective Computing group at MIT for their work developing sensors and software, especially Carson Reynolds and Marc Strauss. We are indebted to

Ken Perlin and John Lippincott for their development of the agent, and to John Rebula for help with system development, data collection and analysis. We would like to thank John D'Auria, Principal of Wellesley Middle School, and their students and staff for being so accommodating of this experiment. We would like to thank Steelcase for loaning us the posture sensor. This material is based upon work supported in part by the Media Lab Things That Think consortium, Deutsche Telekom, by funds from NASA via Ahmed Noor, and by the National Science Foundation under ITR Grant No. 0325428. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation or the official policies, either expressed or implied, of the sponsors or of the United States Government.

References

Allanson, J., Fairclough, S.H., 2004. A research agenda for physiological computing. Interacting with Computers 16 (5), 857–878.

Bickmore, T., 2003. Relational agents: effecting change through human–computer relationships. Ph.D. Thesis, MIT.

Burleson, W., Picard, R.W., 2004. Affective agents: sustaining motivation to learn through failure and a state of stuck. In: Workshop on Social and Emotional Intelligence in Learning Environments.

Burleson, W., Picard, R.W., Perlin, K., Lippincott, J., 2004. A platform for affective agent research. In: Workshop on Empathetic Agents, Third International Joint Conference on Autonomous Agents and Multi-Agent Systems, New York, NY, July 2004.

Card, W.I., Nicholson, M., Crean, G.P., Watkinson, G., Evans, C.R., Wilson, J., Russell, D., 1974. A comparison of doctor and computer interrogation of patients. International Journal of Bio-Medical Computing 5, 175–187.

Chan, T.W., Baskin, A.B., 1988. Studying with the prince: the computer as a learning companion. Montreal, Canada, pp. 194–200.

Conati, C., 2004. Personal conversation. In: Conference on Intelligent Tutoring Systems, August 2004.

Dennerlein, J., Becker, T., Johnson, P., Reynolds, C., Picard, R., 2003. Frustrating computer users increases exposure to physical factors. In: Proceedings of the International Ergonomics Association, Seoul, Korea, August 2003.

D'Mello, S.K., Craig, S.D., Gholson, B., Franklin, S., Picard, R., Graesser, A.C., 2005. Integrating affect sensors in an intelligent tutoring system. In: Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International Conference on Intelligent User Interfaces. ACM Press, New York, pp. 7–13.

Dweck, C., 1999. Self-Theories: Their Role in Motivation, Personality and Development. Psychology Press, Philadelphia.

Fernandez, R., Picard, R.W., 1998. Signal processing for recognition of human frustration. In: International Conference on Acoustics, Speech, and Signal Processing.

Haro, A., Essa, I., Flickner, M., 2000. Detecting and tracking eyes by using their physiological properties. In: Proceedings of Conference on Computer Vision and Pattern Recognition, June 2000.

Herbrich, R., Graepel, T., Campbell, C., 2001. Bayes point machines. Journal of Machine Learning Research 1, 245–279.

Hill, R., Gratch, J., Johnson, W.L., Kyriakakis, C., LaBore, C., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie, J., Rickel, J., Thiébaux, M., Tuch, L., Whitney, R., Douglas, J., Swartout, W., 2001. Toward the holodeck: integrating graphics, sound, character and story. In: Proceedings of the Fifth International Conference on Autonomous Agents, Montreal, Quebec, Canada, pp. 409–416.

Johnson, W.L., Kole, S., Shaw, E., Pain, H., 2003. Socially intelligent learner-agent interaction tactics. In: Proceedings of the 11th International Conference on Artificial Intelligence in Education. IOS Press.

Kapoor, A., Picard, R.W., 2001. A real-time head nod and shake detector. In: Workshop on Perceptive User Interface.

Kapoor, A., Picard, R.W., 2002. Real-time, fully automatic upper facial feature tracking. In: Automatic Face and Gesture Recognition, May 2002.

Kapoor, A., Picard, R.W., 2005. Multimodal affect recognition in learning environments. In: ACM Conference on Multimedia, November 2005.

Kim, H., Ghahramani, Z., 2004. The EM–EP algorithm for Gaussian process classification. In: ECML.

Lepper, M.R., Woolverton, M., Mumme, D.L., Gurtner, J.-L., 1993. Motivational techniques of expert human tutors: lessons for the design of computer-based tutors: lessons for the design of computer-based tutors. In: Lajoie, S.P., Derry, S.J. (Eds.), Computers as Cognitive Tools. Hillsdale, Erlbaum, New Jersey, pp. 75–105.

Lester, J., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A., Bhogal, R.S., 1997. The persona effect: affective impact of animated pedagogical agents. In: Proceedings of Conference on Human Computer Interaction.

Lester, J.C., Towns, S.G., FitzGerald, P.J., 1999. Achieving affective impact: visual emotive communication in lifelike pedagogical agents. The International Journal of Artificial Intelligence in Education 10, 278–291.

Lucas, R.W., Mullen, P.J., Luna, C.B.X., McInroy, D.C., 1977. Psychiatrists and a computer as interrogators of patients with alcohol-related illness: a comparison. British Journal of Psychiatry 131, 160–167.

Malone, T.W., 1984. Heuristics for Designing Enjoyable User Interfaces: Lessons from Computer Games. Ablex Publishing Corporation, Norwood, NJ.

McLaughlin, M., Chen, Y., Park, N., Zhu, W., Yoon, H., 2004. Recognizing user state from haptic data. In: International Conference on Information Systems Analysis and Synthesis.

Mentis, H.M., Gay, G.K., 2002. Using touchpad pressure to detect negative affect. In: IEEE International Conference on Multimodal Interfaces.

Minka, T.P., 2001a. Expectation propagation for approximate Bayesian inference. In: Uncertainty in Artificial Intelligence.

Minka, T.P., 2001b. A family of algorithms for approximate Bayesian inference. Ph.D. Thesis, Massachussetts Institute of Technology.

Moon, Y., 2000. Intimate exchanges: using computers to elicit self-disclosure from consumers. Journal of Consumer Research, 323–339.

Mota, S., Picard, R.W., 2003. Automated posture analysis for detecting learner's interest level. In: Workshop on Computer Vision and Pattern Recognition for Human–Computer Interaction, June 2003.

Picard, R.W., 1997. Affective Computing. MIT Press, Cambridge, MA.

Picard, R.W., 2000. Toward agents that recognize emotion. Vivek 13 (1), 3–13.

Qi, Y., Reynolds, C., Picard, R.W., 2001. The Bayes point machine for computer user frustration detection via pressure mouse. In: Proceedings From the Workshop on Perceptive User Interface.

Reeves, B., Nass, C., 1996. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, New York.

Reynolds, C., 2001. The sensing and measurement of frustration with computers. Master's Thesis, MIT.

Reynolds, C., 2005. Adversarial uses of affective computing and ethical implications. Ph.D. Thesis, September 2005.

Robinson, D., Smith-Lovin, L., 1999. Emotion display as a strategy for identity negotiation. Motivation and Emotion 23 (2).

Robinson, R., West, R., 1992. A comparison of computer and questionnaire methods of history-taking in a genito-urinary clinic. Psychology and Health 6, 77–85.

Rueda, C., 1997. An optimal solution to the towers of hanoi puzzle. URL: ⟨http://yupana.autonoma.edu.co/publicaciones/yupana/003/hanoi/hanoi_eng.html⟩.

Schank, R., Neaman, A., 2001. Motivation and failure in educational systems design. In: Forbus, K., Feltovich, P. (Eds.), Smart Machines in Education. AAAI Press and MIT Press, Cambridge, MA.

Schneider, K., Josephs, I., 1991. The expressive and communicative functions of preschool children's smiles in an achievement situation. Nonverbal Behavior 15.

Strauss, M., Reynolds, C., Hughes, S., Park, K., McDarby, G., Picard, R.W., 2005. The handwave bluetooth skin conductance sensor. In: ACII, pp. 699–706.

Williams, C.K.I., Barber, D., 1998. Bayesian classification with Gaussian processes. IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (12), 1342–1351.