Research and Practice in Technology Enhanced Learning
Vol. 6, No. 2 (2011) 107 – 133

PREDICTING STUDENT EMOTIONS RESULTING FROM APPRAISAL OF ITS FEEDBACK

PAUL SALVADOR INVENTADO

The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, Japan
inventado@ai.sanken.osaka-u.ac.jp

ROBERTO LEGASPI

The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, Japan
roberto@ai.sanken.osaka-u.ac.jp

MERLIN SUAREZ

Center for Empathic Human-Computer Interactions, De La Salle University, 2401 Taft Avenue, Manila,1004, Philippines
merlin.suarez@delasalle.ph

MASAYUKI NUMAO

The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, Japan
numao@ai.sanken.osaka-u.ac.jp

Many researchers have shown the effectiveness of affective ITS for supporting student learning. Support provided to students is usually presented through pedagogical agents capable of expressing emotions through facial expressions, gestures and synthesized speech. Dialogue content is important as it contains information that will help the student learn new information, further understand concepts or correct misconceptions. Although these interventions are based on existing theories, there are still cases when feedback may not fit students as they are very diverse and can be in very different contexts. One very important aspect to consider is how students appraise the feedback given by an ITS. By knowing the student's appraisal of feedback, feedback that is effective and should be retained or replaced can be identified. This research investigates student emotions represented by frustration and excitement values resulting from appraisal of feedback as recognized by an EEG-based device. Frustration and excitement values resulting from feedback appraisal are correlated with feedback from the POOLEIII ITS to create predictive models of these relations. The use of these models will allow future ITS to identify the emotions that result from student appraisal of feedback before it is given, and allow it to perform adjustments on the feedback when necessary.

Keywords: Intelligent tutoring systems; feedback appraisal; affect; brainwave sensors.

1. Overview

Intelligent Tutoring Systems (ITS) are computer systems designed to support students as they learn about a domain by providing assistive feedback (Woolf *et al.*, 2001). Most systems model the student's cognitive state to identify feedback that is appropriate for the student's current context such as giving related quizzes, activities for remedying misconceptions or introducing new topics. Recent works have also made use of affect apart from cognition as it has been shown to play an important role in learning (see Picard *et al.*, 2004 for a noteworthy discussion and survey). Emotions have been utilized in different research such as recognizing affective states while using an ITS (e.g. Kapoor & Picard, 2005; Graesser *et al.*, 2006; Arroyo, Cooper, *et al.*, 2009; Conati & Maclaren, 2009) and providing intervention based on the identified affective states (e.g. Burleson & Picard, 2004; Zakharov, 2007; D'Mello *et al.*, 2008; Woolf *et al.*, 2009).

The feedback provided by ITS is usually created by the designer or an expert as the ITS is developed. Designers identify and map feedback to specific learning cases that students may encounter in the environment that they deem important. Unfortunately, students are very diverse and it is still possible for a student to receive feedback that may not be helpful or even makes the student more confused because of factors that were not considered during the design of feedback. One very important way to identify the appropriateness of feedback is to observe the student's affective reaction toward it. Student reactions toward feedback can then be used to identify when feedback should be retained or adjusted to make them more suitable.

2. Related Works

Intelligent tutoring systems have already been shown to provide learning gains. Most of these ITS however, focus more on the cognitive aspect of student learning but lack the affective aspect which is known to play a significant role in learning (Woolf *et al.*, 2009). Affective ITS have been created to address the affective aspect of learning and include the automatic identification of student emotional states and/or provision of affective intervention to the students' cognitive-affective state (Woolf *et al.*, 2009).

Different approaches have been used in affective ITS for recognizing affect. One approach is inferring affect from student behavior or performance in the ITS such that desirable outcomes can be mapped to positive emotions and failure to reach goals to negative emotions (Zakharov, 2007; Conati & Maclaren, 2009). Another approach is the use of hardware devices like cameras and sensors. Image processing techniques are used to identify facial points from video streams which can be mapped to affective states (Zakharov, 2007; D'Mello *et al.*, 2008; Arroyo, Muldner, *et al.*, 2009). Similarly, data from sensors such as pressure chairs, pressure mouse, and physiological sensors are also mapped to affective states (D'Mello *et al.*, 2008; Arroyo, Muldner, *et al.*, 2009). In both cases, machine learning techniques are usually applied to build models for mapping real time data from these devices to affective states.

Another important aspect of affective ITS is intervention. Using the emotions detected, affective feedback is provided to support student learning. Feedback is usually presented to students using embodied pedagogical agents as they are capable of expressing affect through facial expressions, gestures and synthesized voice (Zakharov, 2007; D'Mello *et al.*, 2008; Arroyo, Muldner, *et al.*, 2009). The content of the dialogue either through text or synthesized speech is very important as well since it provides students with information regarding how they performed, suggestions on how they can solve the problems they are currently facing, motivation, etc. Different theories are used in designing content such as the attribution theory which deals with how students explain their success or failure (Heider, 1958; Weiner, 1985). Providing content which allows students to feel that the reason for their failure can be controlled through effort for example, encourages them to continue. Another theory is cognitive disequilibrium which refers to a state when a person finds his prior knowledge to be incorrect or incomplete after receiving new information (Piaget, 1952). Feedback that moves students into cognitive disequilibrium allow them to learn more. However, allowing them to stay in this state for too long might cause them to disengage or give up (D'Mello *et al.*, 2008).

Some researchers have investigated the effects of feedback to affect. In the work of Pour *et al.* (2010), students were asked to use an ITS in the domain of computer literacy while wearing physiological sensors and the session recorded using a video camera. Students reviewed the video of themselves and reported their affective states when feedback was given. The study found that there was a relationship between feedback and both the self-reported affective and physiological states. In the work of Robison *et al.* (2010) students interacted with an inquiry-based learning environment about microbiology and genetics. While using the ITS, students were asked to identify the appropriateness of the feedback for the current situation and their emotions before and after it was given. Emotion transitions were analyzed relative to the appropriateness of feedback. Their study found that certain emotional states are affected by feedback more compared to others. For example, feedback given while in a state of flow, delight or boredom can easily move the student to a more negative emotional state. This emphasizes the importance of selecting the most appropriate feedback.

This research aims to automatically predict student affective reaction toward feedback. Instead of using discrete emotions, continuous values of frustration and excitement intensities are used. A brainwave sensor that identifies emotions in real time is used to identify student frustration and excitement. This significantly removes the need for manually labeling the data. The domain of the ITS used for this research is for object oriented programming and design which is more complex compared to those used in other research.

3. Feedback Appraisal

It has been shown that student affective states change after receiving feedback from the ITS. What actually causes this change? According to the Appraisal Theory in psychology, emotions are the result of evaluations or appraisals of events that are happening or have happened to a person that causes that person to react (Scherer *et al.*, 2001). The appraisal process consists of different elements such as previous experiences, beliefs and context. However this is complex and there is no research that has completely identified all of these elements. Because of individual differences, these elements differ and the appraisal theory can be used to explain why the same event can cause people to react differently. In the case of ITS, the tutor's feedback can be considered as the event appraised by the student resulting to a reaction. Depending on the student, the same feedback can elicit different reactions which may either help or hinder learning.

The appraisal process would best describe the impact of feedback to students but there is currently no way to observe this process directly. However, appraisal is considered to shape the emotions felt by an individual and different appraisals result in different emotions (Siemer *et al.*, 2007). Thus, by observing the emotion we can approximate the affective direction of the appraisal process. Take the example where a student submits an answer to the system and is given feedback that the answer is incorrect. If a student becomes frustrated, then it is possible that the student thinks the problem is too difficult and he will not be able to solve it. Another student instead may become more engaged indicating he believes the problem is solvable after more effort. Hence, although the entire appraisal process is not understood, looking at the emotion gives a good idea about it. Emotions are manifested in different ways such as facial expressions, speech, gestures and physiology which can be used for identifying them. In this research, emotion states are identified automatically using the Emotiv EPOC Neuroheadset, a commercially available electroencephalogram (EEG) based device to measure the student's emotion while using an ITS. The emotion identified by the device after feedback is given to a student is used to approximate the student's appraisal.

4. Emotiv EPOC Neuroheadset

The Emotiv EPOC Neuroheadset [a] collects brainwave signals from 14 sensors touching the scalp. It provides a standard developer kit (SDK) which extracts frustration and excitement intensities from brain signals in real time. These intensities are represented as values ranging from 0.0 to 1.0. According to the developers, emotion models were built using data gathered from a group of over a hundred volunteers. Different activities such as gameplay, watching videos and psychometric photography sets were used to induce emotions. Videos of the subjects were recorded and data from physiological sensors for respiration, skin conductance and heart rate and blood volume flow were used together with EEG data collected using the Emotiv EPOC. Finally, subjects answered questionnaires about their experience. All data were analyzed and used to help psychologists label the data with emotion labels. Features from the data that was gathered and known to be specific to particular emotions were used in the creation of the emotion models. When the emotion models are used in real-time, the system performs self-scaling

[a] http://www.emotiv.com

such that it is able to adapt to the base point and range of emotions of the current user. Figure 1 shows an image of a student wearing the Emotiv EPOC Neuroheadset.

Figure 1. Student wearing the Emotiv EPOC Neuroheadset.

5. POOLE III ITS

The Programmer's Object Oriented Learning Environment III (POOLE III) is an ITS designed to support students learning object oriented programming (Chan *et al.*, 2004; Aurellano *et al.*, 2007). The version used for this research focuses specifically on the creation of class diagrams using the Unified Modeling Language (UML). The system includes a virtual human agent that provides feedback to the student while learning in the environment. The agent is capable of expressing facial expressions and synthesized speech. Figure 2 shows a screenshot of the ITS and the virtual agent.

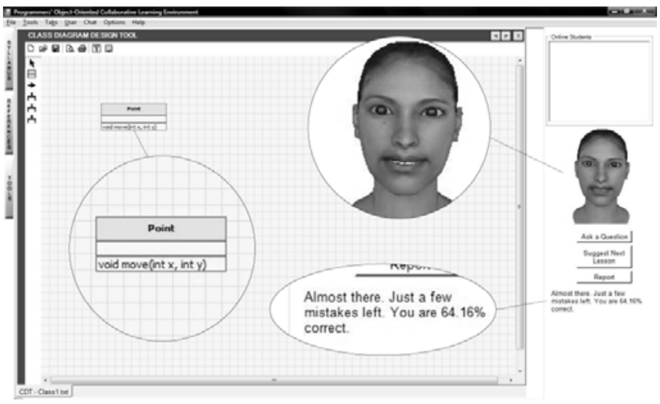

Figure 2. Screenshot of POOLE III.

The student's cognitive model is represented using a Bayesian network of the different topics discussed in the ITS and is used as basis when providing suggestive feedback to the student. As the student interacts with the ITS, it provides support to the student through three kinds of feedback: activity transition, solution evaluation and hints. Activity transition feedback is given when the student moves to a different activity within the system such as viewing a new lesson or completing a different exercise. Solution evaluation feedback is given when students ask the system to check their solution for a given problem. The correctness of the student's solution is identified using an intention based analysis. The student's solution is matched to the most similar expert solution stored in the system and the differences from the matched expert solution are used to grade the solution. Lastly, hints are provided upon request by the student and provide suggestions on what they can do to correct their work. The hints are designed to become more specific as students ask for more hints. This allows the student to have the opportunity to get more insights about the problem and solve it independently. Students can continuously ask for more hints if they find themselves stuck on the problem.

The affective aspect of POOLE III lies in its intervention mechanism. The dialogue for the feedback was designed to encourage students to continue learning by providing praise when the student does well and providing reassurance despite the difficulty experienced while solving the problem. There are a total of ten activity transition messages, five solution evaluation messages and eight hint messages giving a total of 23 message templates. However, these messages change depending on the content such as the name of the topic or lesson for activity transition messages, grade of the student for solution evaluation messages and the focus of the hint for hint messages. Table 1 shows a subset of the feedback provided for the student and a description of the cases when these dialogues are used.

Table 1. POOLE III Feedback.

Type	Case	Feedback
Activity Transition	Student requests a correct solution	"It would be great if you can identify the mistakes yourself next time."
	Student continuously skips lessons	"Maybe you should try solving the current problem first before jumping to the next one."
Solution Evaluation	Student submits the correct solution on the first try	"Wow! You got this on your first attempt! See if you can solve another problem just as fast!"
	Student submits an incorrect solution with a few errors	"Almost there, there are a few mistakes left. You are <x>% correct."
Hint	Student gets a hint for the 1st time	"Try recalling your lectures for this lesson."
	Student gets a hint for the 8th time	"The <mistake> is causing the problem."

6. Data Gathering

Data was gathered from ten first-year college students taking an introductory course on object oriented programming. Five of the students were male and five were female selected randomly from 120 students of the entire batch. Each student attended a single session that lasted for 20 minutes. Students were in a closed room with a laptop and a video camera focused on them. No one else was allowed to enter the room to avoid any disruption or distraction that will result to noises in the data.

At the start of the session, the student was given a short tutorial on using POOLE III. The student was then asked to wear the Emotiv EPOC and a video camera was focused on the student while using POOLE III. The video images were later used to identify the start and end of each appraisal. During the session, the frustration and excitement values from the Emotiv EPOC were stored together with the logs of the feedback from POOLE III. After the session, students were asked to answer a survey about their experience in using POOLE III and to provide their student profile. They were also asked to answer a Big Five Personality test (Goldberg, 1993) to identify their levels of accommodation (how willing individuals are to assist others), emotional stability (how varied a person's emotions are), extroversion (how comfortable a person is to interact with others), inquisitiveness (how curious a person is in learning or understanding something) and orderliness (how organized a person's thoughts or activities are). In each dimension, a person is scored on how much this dimension is seen in the person's behavior. Unlike other personality tests that classify a person to fall under a finite set of categories, personality is described in terms of levels of the five dimensions allowing a broader spectrum of personality. The level of these dimensions relate to how a person behaves which we feel can be used to account for how a student would react to feedback.

7. Data Preprocessing

Our research focused on feedback appraisal and therefore we only utilized the data from Emotiv EPOC representing states immediately before feedback was given and states when the appraisal process was finished. As there is no study indicating the duration of time when prior emotions affect appraisal, initially all instances 20 seconds prior to giving feedback were retained.

Based on the earlier discussions, we consider that the appraisal process results in an emotion prompting a person to perform an action. In the case of our experiment, we consider the student's reaction to be the point when the student resumes using POOLE III after receiving feedback. Thus, the appraisal process has completed at this point, and the emotion from the appraisal is considered to be the frustration and excitement values obtained from the Emotiv EPOC at that point. The specific time that this happens can easily be taken from the logs of the ITS and verified manually by viewing the videos taken of the student using the system.

All these values are then combined to form the instances used in the machine learning algorithms discussed in the next section.

Table 2 shows the list of all features and labels used in the experiment. There were two models created using machine learning. First is the excitement model that was built using the different features and the post-appraisal excitement labels. Second is the frustration model that was created with the post-appraisal frustration label. Figure 3 shows three sample instances that use these features and the post-appraisal excitement label.

Table 2. List of features and labels.

Feature	**Description**
age	Age of the student
gender	Gender of the student (0 = male, 1 = female)
accommodation	Values of the elements from the Big Five personality test taken by the student
emotional_stability	
extroversion	
inquisitiveness	
orderliness	
feedback	Feedback message given to the student
excitement	Excitement value when feedback is given
frustration	Frustration value when feedback is given
ave_excitement$_{10}$	Average of excitement values ten seconds prior to giving feedback
ave_excitement$_{5}$	Average of excitement values five seconds prior to giving feedback
ave_excitement$_{2}$	Average of excitement values two seconds prior to giving feedback
std_deviation_excitement$_{10}$	Standard deviation of excitement values ten seconds prior to giving feedback
std_deviation_excitement$_{5}$	Standard deviation of excitement values five seconds prior to giving feedback
std_deviation_excitement$_{2}$	Standard deviation of excitement values two seconds prior to giving feedback
ave_frustration$_{10}$	Average of frustration values ten seconds prior to giving feedback
ave_frustration$_{5}$	Average of frustration values five seconds prior to giving feedback
ave_frustration$_{2}$	Average of frustration values two seconds prior to giving feedback
std_deviation_frustration$_{10}$	Standard deviation of frustration values ten seconds prior to giving feedback
std_deviation_frustration$_{5}$	Standard deviation of frustration values five seconds prior to giving feedback
std_deviation_frustration$_{2}$	Standard deviation of frustration values two seconds prior to giving feedback
Label	**Description**
post_appraisal_excitement	Excitement value after appraisal
post_appraisal_frustration	Frustration value after appraisal

There were a total of 235 feedback messages given by the system to the ten students which were used as the instances for the machine learning algorithm. Each instance is represented as a vector composed of the 22 features and the frustration or excitement label. Out of the 235 feedback given to all the students, 14.89% were about activity transitions, 43.40% were about hints and 41.70% were about assessment.

```
<17,0,82,66,46,80,62,"There are a couple of    ...",0.360,0.610,0.512,0.424,0.385,0.101,0.036,0.007,0.632,0.532,0.515,0.121,0.051,0.078,0.390>
<17,1,66,66,40,42,70, "There are a couple of   ...",0.042,0.488,0.109,0.054,0.034,0.065,0.039,0.032,0.660,0.616,0.583,0.057,0.040,0.048,0.179>
<18,0,52,72,32,52,68," You seem to be having ...",0.626,0.172,0.506,0.401,0.383,0.151,0.040,0.039,0.357,0.285,0.240,0.090,0.044,0.034,0.491>
```

Figure 3. Sample instances of preprocessed data. The last element in the vector represents the excitement value after feedback appraisal.

8. Machine Learning

Since two emotions are measured using the Emotiv EPOC, two models have to be created to predict the student's level of frustration and excitement after feedback is given. The data can be viewed as two sets, where both have the same instances but one set uses the frustration label and the other the excitement label. On both data sets, feature selection is applied to identify the most significant features. Specifically, the forward feature selection algorithm (Mierswa *et al.*, 2006) was used together with a linear regression, k-nearest neighbor or support vector machine wrapper. The resulting data with feature subsets selected from using a linear regression wrapper was used as the training data for creating the model with linear regression. Similarly, the data with feature subsets selected from using the k-nearest neighbor wrapper was used as training data for creating the k-nearest neighbor and the support vector machine wrapper for the support vector machine model.

Feature selection reveals which subset of the existing features in the test bed will ultimately be viable for prediction. Fewer features that result from removing irrelevant and redundant features improve concept learning in terms of speed and concept generalization. Although feature selection consequently creates a new search space with lower dimensions, it still preserves the bases for analyses and interpretations of the data and its structure (Legaspi *et al.*, 2010). Feature selection requires a candidate feature subset search strategy and an objective function to evaluate the candidate subset. The forward search starts with a null hypothesis and then adds one feature at a time that will maximize the objective function when combined with the selected features. As an objective function, wrappers evaluate the candidate subsets based on how accurate a learning algorithm can approximate the concept to be generalized using statistical resampling or cross validation techniques (Legaspi *et al.*, 2010).

For this research, supervised machine learning tasks were performed. Figure 4 illustrates the learning framework for constructing each of the pertinent models. The first task was the creation of models for predicting student frustration and excitement intensities after receiving feedback. As stated earlier, the labels used are taken from the Emotiv EPOC. The second task was the creation of a model for predicting the amount of increase or decrease in frustration and excitement after feedback is given. The labels used in this task were the difference between the frustration and excitement values before and after the feedback was given. Both tasks create predictive models, thus, the features used only describe the instances before feedback is given. The labels used are continuous values so regression machine learning algorithms were used. Specifically, the machine learning algorithms used were linear regression, k-nearest neighbor (k=1...3) and support vector machine using a dot product kernel. For the k-nearest neighbor, initial models were created with varying values for k. When values above four were used, there were already very small increases and decreases in performance.

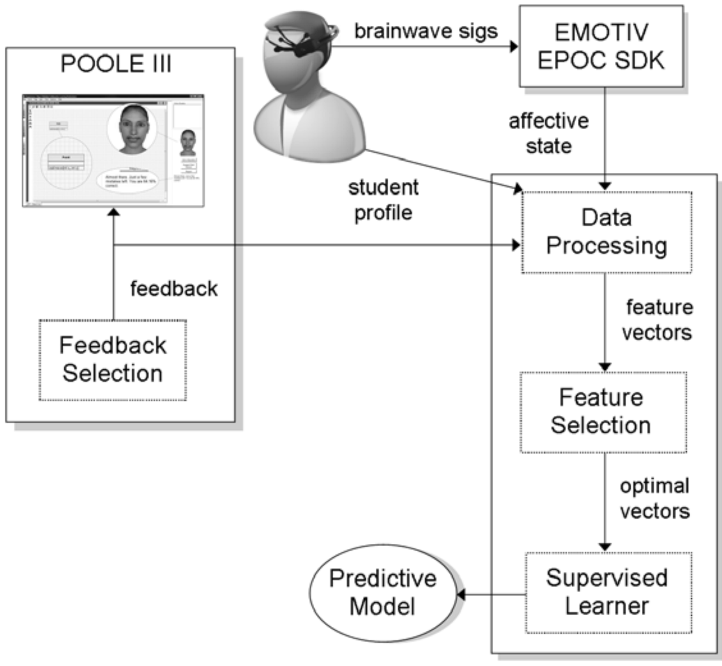

Figure 4. Machine learning framework.

Batched cross-validation (Mierswa *et al.*, 2006) was used for evaluating the models. Batched cross validation differs from cross-validation in the way instances are split as training and test data. In batched cross-validation, all instances belonging to a batch are ensured to be either in the testing or training set. No instances belonging to a batch are split. In the case of the experiment, the student ID is used to specify a batch, hence, one batch being all the instances associated to one student. In effect, the behavior of a single student is either learned or tested with the behaviors of all the other students. This is a better test of the generality of the model created compared to treating instances of students independently.

8.1. *Linear regression*

Linear regression is a method in statistics which is also used as a machine learning technique for creating a linear model that will approximate the class of unseen data based on a given set of training examples (Witten & Frank, 2005). The linear regression model is represented as a linear combination of weights and feature values:

$$x = w_0 + w_1 a_1 + w_2 a_2 + \ldots + w_k a_k \,. \tag{8.1}$$

In Eq. (8.1), x refers to the class; $a_1, a_2, \ldots, a_k$ are the feature values; and $w_0, w_1, \ldots, w_k$ are the weights. In a regression problem the class x is a continuous value computed by the sum of the products of the feature and weight values. For the equation to model the training examples correctly, weight values have to be identified such that for each example, the predicted class resulting from applying the equation on its features is the same as its actual class. In most cases, the actual target values cannot be modeled perfectly, so the model with the least error is considered to be the best fit model. Error is usually computed by solving for the total squared error between the actual and predicted classes of all training examples:

$$\sum_{i=1}^{n}\left(x^{(i)}-\sum_{j=0}^{k} w_j a_j^{(i)}\right)^2 . \tag{8.2}$$

In Eq. (8.2), $x^{(i)}$ refers to the target value and the summation function computes the predicted value by computing the linear combination of each weight value, w_j, and feature value $a_j^{(i)}$. The best fit model is identified by adjusting the weight values and minimizing the squared error by using minimization techniques such as gradient descent. Gradient descent is an iterative process which updates weight values based on the difference between the target value and predicted value:

$$w_i' = w_i + \alpha(x^{(i)} - p^{(i)})a^{(i)} . \tag{8.3}$$

In Eq. (8.3), w_i' refers to the updated weight value, w_i refers to the current weight value, $x^{(i)}$ refers to the target value, $p^{(i)}$ refers to the predicted value using the current set of weight values and $a^{(i)}$ refers to the feature value. Weight values are continuously updated until certain thresholds are met. It may be based on the amount of error, change in error, changes in weight values or other factors. Once the best fit model is identified, the features of unseen data are plugged into the equation to predict its class value.

8.2. *k-Nearest neighbor*

k-Nearest neighbor is an instance based machine learning algorithm which identifies the class of unseen data based on the most similar example from the training set (Witten & Frank, 2005). It is considered a "lazy" algorithm because it searches the training set for the most similar instance at the time when classification or regression is performed. The simplest way of implementing the algorithm is applying a distance function such as the Euclidean distance formula to compute the distance between the features of the unseen data and those of each example in the training set. The value *k* defines how many of the most similar instances or, nearest neighbors, is considered for identifying the unseen data's class. For a regression problem, when *k* is one, the numeric class of the nearest neighbor is used as the predicted class for the unseen data. When *k* is greater than one, the distance-weighted average of the nearest neighbors' numeric classes are used as the predicted class of the unseen data:

$$p = \frac{\sum_{i=0}^{k}\frac{1}{d_i^2}c_i}{\sum_{i=0}^{k}\frac{1}{d_i^2}} . \tag{8.4}$$

In Eq. (8.4), p is the predicted class value, d_i is the distance of the i^{th} neighbor and c_i is the class value of the i^{th} neighbor.

8.3. *Support vector regression*

Support vector machine (SVM) (Vapnik, 1999) is a machine learning algorithm used for classification. However, it can also be applied to regression problems and is also referred to as support vector regression (SVR) (Smola & Schölkopf, 2004). The goal of SVR, just like linear regression, is to find a linear function that would fit the given training data so that the resulting function can be used to predict the target value of unseen data. Linear functions take the form:

$$f(x) = w^T x + b\,. \tag{8.5}$$

In Eq. (8.5), w^T refers to the weight vector, x refers to the input vector and b refers to the bias. The goal is to find a function that is as flat as possible and can make predictions having at most ε deviation from the target value. Flatness refers to finding a small value for w. This can be achieved by minimizing the norm and can be written as a convex optimization problem:

$$\begin{array}{ll} \textit{minimize} & \frac{1}{2}\|w\|^2 \\ \textit{subject to} & \begin{cases} y_i - \langle w, x_i \rangle - b & \leq \quad \varepsilon \\ \langle w, x_i \rangle + b - y_i & \leq \quad \varepsilon \end{cases} \end{array}. \tag{8.6}$$

In Eq. (8.6), y_i refers to the target value and ε is a user specified threshold value which controls how fit the function will be to the training data. However, there are cases when no such function can be found so some amount of error is allowed by introducing slack variables ξ_i, ξ_i^* . This results to:

$$\begin{array}{ll} \textit{minimize} & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ \textit{subject to} & \begin{cases} y - \langle w, x_i \rangle - b & \leq \quad \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i & \leq \quad \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq \quad 0 \end{cases} \end{array}. \tag{8.7}$$

In Eq. (8.7), the constant $C>0$ determines the trade-off between flatness and the deviations from ε. Lagrange multipliers α_i^* and α_i are introduced resulting in a regression estimate of the form:

$$f(x) = \sum_{i=1}^{m}(\alpha_i^* - \alpha_i)\langle x_i, x \rangle + b\,. \tag{8.8}$$

Equation (8.8) is called the Support Vector Expansion where w is described by a linear combination of training patterns x_i. Both SVM and SVR can create models that handle non-linearly separable data by mapping features of the training data to higher dimensions. In this higher dimension, linear functions can be constructed which when mapped back to the original feature space, creates non-linear boundaries. Since the support vector expansion depends on dot products between training patterns, these can be substituted by functions that satisfy Mercer's condition and map them to higher dimensional spaces. These functions are called kernel functions. Given a kernel function k, the support vector expansion can be rewritten as:

$$f(x) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) k(x_i, x) + b \, . \tag{8.9}$$

The use of kernels makes computation without the need to worry about the increase in dimensionality. Different kernels produce different feature spaces and thus different prediction values. The resulting function configuration can then be used to predict the target value of unseen data.

9. Results and Analyses

The feature selection and machine learning tasks were performed using the RapidMiner data mining software (Mierswa *et al.*, 2006). While the performance of the machine learning models was analyzed, the features related to prior emotional values of frustration and excitement were also tested. Initially, the average and standard deviation of frustration and excitement values from the previous 20 seconds were included in the feature set. As this duration was increased, there were no improvements in performance. However, when this value was decreased, the best performance was found when the average and standard deviation of frustration and excitement from the previous ten seconds were used in the feature set. The results of analyzing the data and the machine learning tasks are discussed in the following subsections.

9.1. *Feedback and affect relations*

Assessment feedback provides students with scores for their answers. We feel that among the feedback types, assessment elicits emotions more since it affects the student's goals in studying, which in this case would be getting the right answer. Figure 5 shows the changes in frustration of students when they receive feedback. The change in frustration was measured by using the formula $\Delta emo_i = emo_{aa} - emo_{ba}$ where emo_{aa} refers to the emotion after appraisal and $emotion_{ba}$ refers to the emotion before appraisal. Positive values indicate increase in frustration while negative values indicate decrease in frustration. As shown in the figure, when students were presented with higher scores (i.e. 64.16% to 86.56%) they experienced more frustration. This may seem counterintuitive but further investigation of the videos and system logs showed that in most instances students spent most of the time being stuck at these points. They changed many elements in the class diagram but it usually results in them having the same or even lower scores. This state of being stuck for a prolonged time would be the probable cause for frustration. Although there was more frustration when students got higher scores the nearer the scores were to the perfect answer, students showed less frustration probably because of their feeling of being on the right track and nearing the completion of the exercise. Interestingly, students still experienced frustration when getting the correct answer. This is probably because, of the few students who got the correct answers, it was observed in the video and informal interviews that even though they got the right answer they realized that they simply overlooked certain elements in their solution causing their answer to be incorrect after spending a lot of time trying to solve it, making it frustrating for them.

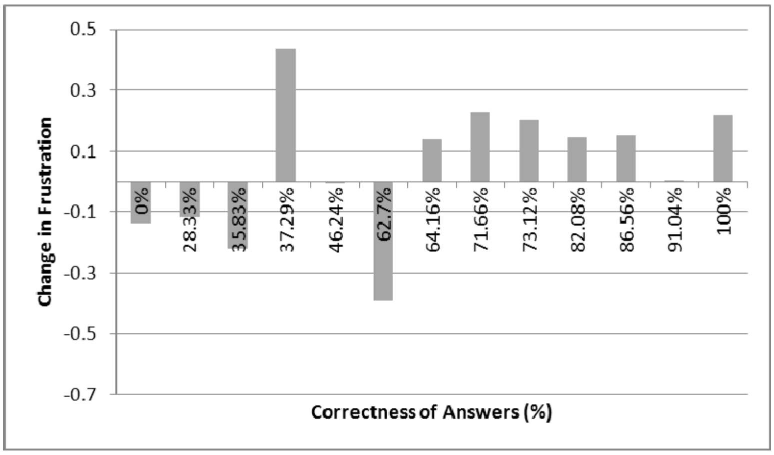

Figure 5. Change in frustration levels of students when given feedback about the correctness of their answers. Positive values indicate increase in frustration while negative values indicate decrease in frustration.

When students got lower scores, frustration was also lower since there was probably less pressure as students feel they were still far from getting the correct answer. Slight modifications of their answers at this point also provided big improvements to their scores since there were more elements they explored and manipulated.

Figure 6 there were bigger changes in the excitement values when students' scores are lower. Around this time, students were observed to explore more as there were more elements that they could manipulate. This exploration and discovery of configurations that improved their scores could have easily caused excitement. As discussed earlier, it was found that students usually felt stuck when they had higher scores. Being in a state of stuck, students were not able to experience much excitement as they would probably just try very minute changes to their original solutions.

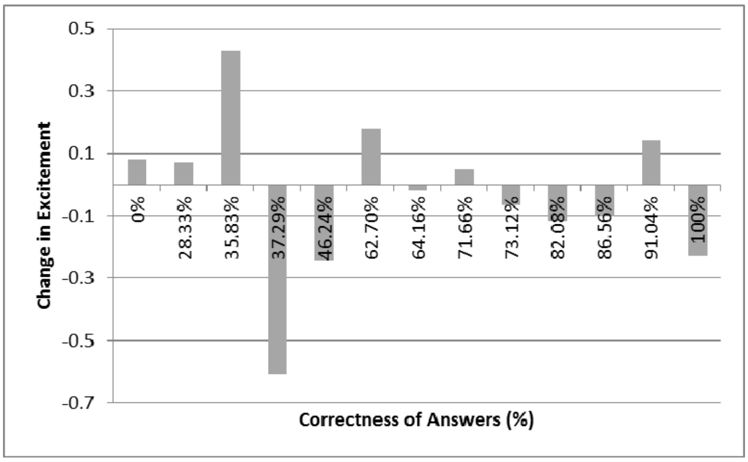

Figure 6. Change in excitement levels of students when given feedback about the correctness of their answers. Positive values indicate increase in excitement while negative values indicate decrease in excitement.

The observations here are important since they indicate how assessment feedback can be designed. For example when scores are lower, students will probably be more open to suggestions and criticism as there is lesser frustration and more excitement. However, the student might need more support and encouragement when they stay too long in the problem. More encouragement would probably help students experiencing lower excitement intensities by helping them to be more involved in the activity. Once they are more excited about the activity, feedback has more impact on the student and will probably help them learn better.

Whenever students have a hard time finding the answer, they can ask for hints from the system. Hints were generated using 8-hint templates. As the student asked for more hints, the hint became more and more specific. For example, the first time a student asked for a hint the system asked the student to review his notes first. As the student asked for more hints the student was given a concept in object oriented programming involved in the error (e.g. “There seems to be something wrong with your attributes.”). The most specific hint given to the student was the actual element causing the problem (e.g. “The missing attribute age is causing the problem.”). Figures 7 and 8 show the changes in frustration and excitement of students when they received hints from the system. The changes in frustration and excitement are calculated using the same formula as that of the changes in emotions for assessment feedback, which is $\Delta emo_i = emo_{aa} - emo_{ba}$. From the results, the hints are observed to decrease frustration and increase the excitement of the students. Decrease in frustration may be caused by students realizing what they are doing wrong or that they have a chance to get the correct answer by following the hint. Increase in excitement may be drawn from feelings of seeing if applying what they learn from the hint will allow them to get the correct answer. Although some hints resulted in more changes in frustration and excitement than others, it is not enough to conclude that certain types of hints are better than others. The use of hints is very contextual such that students may consider a certain type of hint very useful at a particular situation but less in another.

The correlation between the number of times students asking for hints from the system and changes in their frustration and excitement are not significant having the values 0.009 and 0.036, respectively. This indicates that it is not the number of times hints were asked that caused changes in the emotion. The appropriateness of the content will still probably bear more weight on frustration and excitement levels.

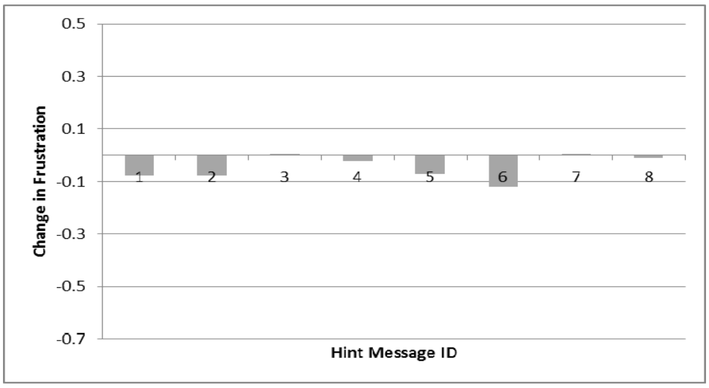

Figure 7. Change in frustration levels of students after receiving hints.

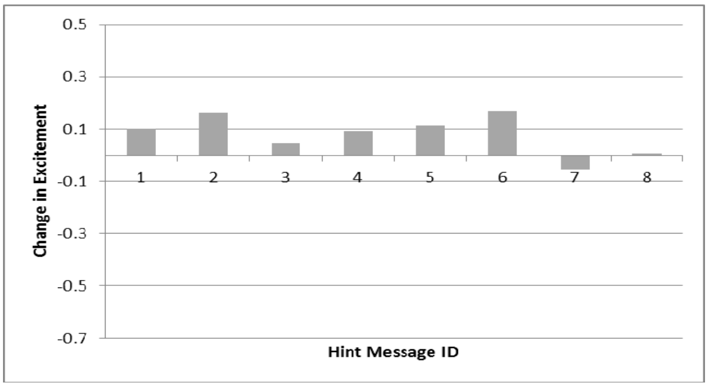

Figure 8. Change in excitement levels of students after receiving hints.

When students finished solving a problem or decide to move to a different lesson or problem without finishing the current one, students received transition feedback. There were four transition feedback given by the system as shown in Table 3.

Table 3. Transition messages given by the system.

ID	Transition message
1	Okay let's try learning this next lesson. <Lesson title>.
2	Okay let's try doing this next problem. <Problem name>.
3	Maybe you should try solving the current problem first before jumping to the next one.
4	It would be great if you can identify the mistakes yourself next time.

Student frustration and excitement values when receiving the transition messages are shown in Figures 9 and 10. It is important to note that only two out of ten students were able to solve one problem each. Apart from these two instances, all other transitions were carried out without completing an exercise.

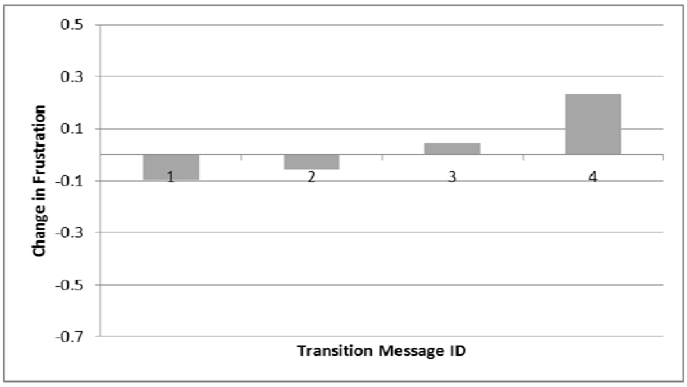

Figure 9. Changes in frustration values when students receive transition feedback. The ID of the transition message is used to label the corresponding frustration value.

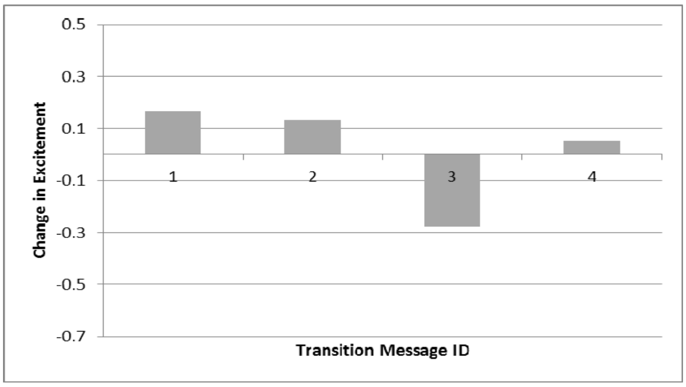

Figure 10. Changes in excitement values when students receive transition feedback.

Interestingly for transition messages 1 and 2, there was a decrease in frustration and increase in excitement. These feedback was given when the student moved to a different problem or lesson altogether. Decrease in frustration was most likely caused by the decrease in feelings of pressure as the student no longer worried about the previous problem being solved. Increase in excitement on the other hand was most likely caused by the new activity that the student would engage in. It was a chance for the student to try again and have the chance to get the answer correctly. For transition message 3, the student was told to continue solving the current problem which they had not yet solved instead of solving a new one. Students showed increase in frustration and decrease in excitement. Increase in frustration may be caused by the hindrance to the student's goal of moving into the next problem. When the student decided to move to the next problem, the student would already be looking forward to solving the next problem and stopping thinking of the current problem. However, by not allowing the student to do so, the student needed to settle solving the current problem again as well which could be causing the increase in frustration and decrease in excitement. For transition message 4, the student was told that more effort was needed on the student's part. Both frustration and excitement were observed to increase. This showed that the students interpreted this feedback with negative emotion probably because it sounds less supportive relative to the effort that was put in by the student. At the same time, the certain degree of excitement may have been brought by the student being challenged to do better next time.

9.2. *Predicting student frustration and excitement after feedback*

From an initial set of 22 features, the most number of features in a single data set was reduced to ten, which is about 55% reduction of the original features. For all machine learning algorithms, the use of feature selected data resulted in significant improvements in accuracy over the original set of features. Tables 4 and 5 show the features selected by the feature selection algorithm using their respective machine learning wrappers and the percentage of features reduced from the original set of features. Each feature was selected depending on their contribution towards an optimal prediction performance of the learning algorithm. It is therefore the case that the feature subset found to be optimal for one learning algorithm can prove detrimental to the performance of another algorithm. It is also the case that the number of features selected does not affect the prediction performance of the algorithm (e.g. having fewer features does not equate to less accurate predictions). For example, as will be shown later on, despite of having only one feature selected for the SVM (in Table 4), it still outperformed the k-NN algorithms. What will be made clear later on here is that the selected feature subsets for both the frustration and excitement models led to improvements in results.

Table 4. Selected features for the Frustration Model.

Feature	LR	1-NN	2-NN	3-NN	SVM
Age					
Gender					
Accommodation	O				
emotional_stability					
Extroversion	O				
Inquisitiveness	O				
Orderliness					
Feedback	O				
Excitement	O		O	O	O
Frustration	O				
$ave\_excitement_{10}$		O			
$ave\_excitement_{5}$					
$ave\_excitement_{2}$	O				
$std\_deviation\_excitement_{10}$		`		O	
$std\_deviation\_excitement_{5}$	O			O	
$std\_deviation\_excitement_{2}$	O		O		
$ave\_frustration_{10}$					
$ave\_frustration_{5}$					
$ave\_frustration_{2}$					
$std\_deviation\_frustration_{10}$	O				
$std\_deviation\_frustration_{5}$					
$std\_deviation\_frustration_{2}$					
Reduction	54.55%	95.46%	90.91%	86.36%	95.46%

Table 5. Selected features for the Excitement Model.

Feature	LR	1-NN	2-NN	3-NN	SVM
Age	O				
Gender					
Accommodation					
emotional stability					O
Extroversion	O		O		
Inquisitiveness	O				
Orderliness					
Feedback					
Excitement					
Frustration	O	O	O	O	
$ave\_excitement_{10}$					O
$ave\_excitement_{5}$			O	O	O
$ave\_excitement_{2}$	O				
$std\_deviation\_excitement_{10}$					
$std\_deviation\_excitement_{5}$	O				
$std\_deviation\_excitement_{2}$		O	O	O	
$ave\_frustration_{10}$					
$ave\_frustration_{5}$			O		
$aver\_frustration_{2}$	O				
$std\_deviation\_frustration_{10}$	O				
$std\_deviation\_frustration_{5}$	O				
$std\_deviation\_frustration_{2}$	O		O		
Reduction	54.55%	90.91%	77.27%	86.36%	86.36%

It is also interesting to note that only the frustration model uses feedback as a feature. Looking at the relations of frustration and excitement to the appraisal of feedback, for transition and hint feedback both frustration and excitement were affected by the feedback and there was an inversely proportional relationship between the two. However, in assessment feedback only the changes in frustration were observed to be affected by the content of the feedback. For excitement values, its relationship between feedback was not as clear and may be a reason why feedback was not used as a feature. It is probable that the relationship between frustration and excitement was stronger than the feedback for the excitement model and the reason why frustration related features were retained and feedback was removed. We need to mention again that it is how the student appraises the feedback and not the feedback per se that influences the student's affective reactions. Hence, it is still possible that feedback has no correlation with the emotion classes but other factors like the student's personality and the affective states prior to receiving the feedback that shape the emotional responses.

The performances of the models created by the learning algorithms were measured using the coefficient of determination (R^2). R^2 measures how well a model can predict future outcomes based on existing data, which in this case is the affective data collected through the Emotiv EPOC. Since correlation shows how fit the model is to the existing data, it also measures how much the predictions follow the behavior of the target values (i.e. model predicts high values for high target values and low values for low target values). Student behaviors may be inferred depending on these values. For example high frustration values may indicate how negatively the student sees the feedback or high excitement can indicate how interested the student is toward the feedback.

As shown in Figures 11 and 12, linear regression gave the best results for creating predictive models of frustration and excitement. Using linear regression on the feature selected frustration dataset resulted in a model with an R^2 of 0.462 and the feature selected excitement dataset resulted in a model with an R^2 of 0.560. The linear regression models for frustration excitement are shown in Eqs. (9.1) and (9.2) respectively.

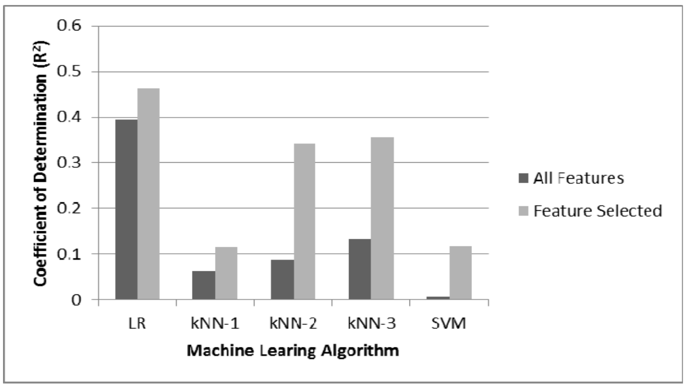

Figure 11. Coefficient of determination (R^2) values of resulting machine learning models for predicting frustration with and without feature selection.

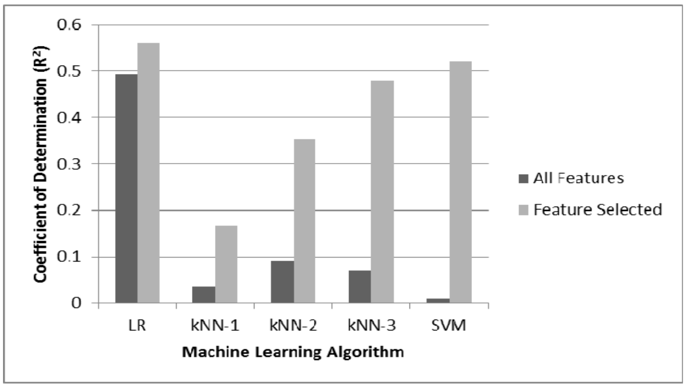

Figure 12. Coefficient of determination (R^2) values of resulting machine learning models for predicting excitement with and without feature selection.

0.001 * *Feedback* + 0.823 * *Excitement* - 0.003 * *Extroversion* + 0.615 * *SDEx2* + 0.176 * *Frustration* - 0.334 * *SDFr10* + 0.004 * *Inquisitiveness* - 0.003 * *Accommodation* - 0.169 * *AveEx2* - 0.331 * *SDEx5* + 0.184 (9.1)

1.111 * *Frustration* - 0.003 * *Extroversion* - 0.451 * *AveFR2*- 0.232 * *SDFR5*- 0.002 * *Inquisitiveness* + 0.085 * *AveEx2* - 0.018 * *Age* - 0.266 * *SDEx5* - 0.254 * *SDFR2* - 0.136 * *SDFr10* + 0.729 (9.2)

Linear regression may have performed better compared to the other algorithms because of the weights it attributes to each feature value. Features that have more contribution to the resulting excitement or frustration intensity are given more weight while those that are not as relevant or are noisy are given smaller weights. These results in predictions based more on the features that actually contribute to the prediction. k-NN on the other hand predicts using the similarity of the features of new data with existing data. When calculating distance, noisy features easily increase distance. Since brainwave data is continuous and tends to be noisy, the predictions made by the k-NN model were not as good. The results of SVM are dependent on the kernel used. It is possible that the dot product kernel, which was used for the experiment, was not able to find the best fit for the data thus resulting in poorer prediction.

9.3. *Predicting change in student frustration and excitement after feedback*

Feature selection in this machine learning task also resulted in around 55% reduction of the number of features. Apart from SVM, the use of feature selected data resulted in improvements with the machine learning models created as shown in Figures 13 and 14. Tables 6 and 7 show the features selected by the feature selection algorithm using their respective machine learning algorithm wrappers and the percentage of features reduced from the original set of features. There are differences in the features selected in this machine learning task compared to the previous machine learning task. It is important to note however, that features related to previous frustration and excitement values were also selected like the previous machine learning task. This may indicate that there is a relationship between the affective states before feedback is given and the resulting affective state after feedback is given.

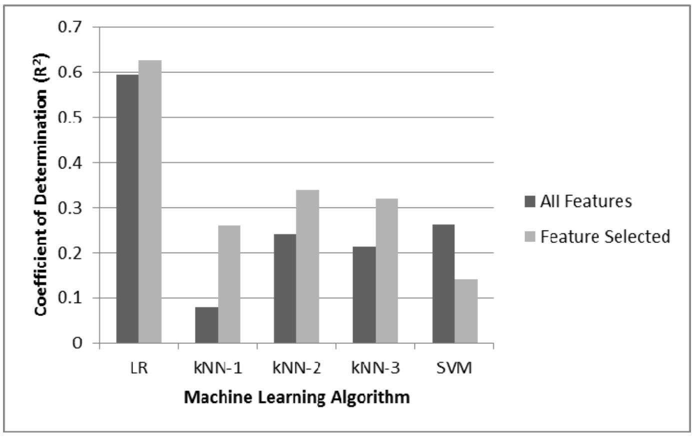

Figure 13. Coefficient of determination (R^2) values of resulting machine learning models for predicting change in frustration with and without feature selection.

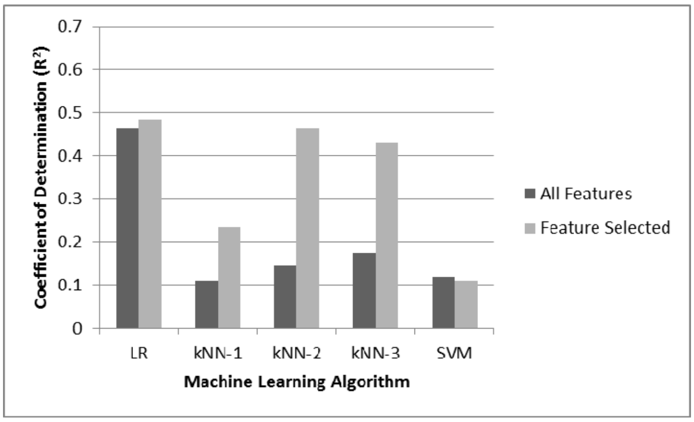

Figure 14. Coefficient of determination (R^2) values of resulting machine learning models for predicting change in excitement with and without feature selection.

Table 6. Selected features for the change in Frustration Model.

Feature	LR	1-NN	2-NN	3-NN	SVM
age					
gender	O				O
accommodation	O				
emotionl_stability					
extroversion	O			O	
inquisitiveness	O				
orderliness					
Feedback	O				
excitement	O		O	O	
frustration	O		O		
ave_excitement$_{10}$		O			
ave_excitement$_{5}$					
ave_excitement$_{2}$	O				
std_deviation_excitement$_{10}$					
std_deviation_excitement$_{5}$					
std_deviation_excitement$_{2}$	O		O		
ave_frustration$_{10}$				O	
ave_frustration$_{5}$					
ave_frustration$_{2}$				O	
std_deviation_frustration$_{10}$	O			O	
std_deviation_frustration$_{5}$					O
std_deviation_frustration$_{2}$				O	O
Reduction	54.55%	95.46%	86.36%	72.72%	86.36%

Table 7. Selected features for the change in Excitement Model.

Feature	LR	1-NN	2-NN	3-NN	SVM
age					
gender	O				O
accommodation					
emotional_stability					
extroversion	O				
inquisitiveness	O				
orderliness					
feedback					
excitement	O		O	O	
frustration	O		O	O	
ave_excitement$_{10}$		O			
ave_excitement$_{5}$					
ave_excitement$_{2}$					
std_deviation_excitement$_{10}$					
std_deviation_excitement$_{5}$	O			O	
std_deviation_excitement$_{2}$	O		O		
ave_frustration$_{10}$					
ave_frustration$_{5}$					
ave_frustration$_{2}$	O		O	O	
std_deviation_frustration$_{10}$	O				
std_deviation_frustration$_{5}$	O				
std_deviation_frustration$_{2}$				O	
Reduction	54.55%	95.46%	81.81%	77.27%	95.46%

The performances of the models were again measured using the coefficient of determination. The frustration model had an R^2 of 0.627 and the excitement model had an R^2 of 0.484. The linear regression model for the frustration model is shown in Eq. (9.3) and the excitement model in Eq. (9.4). Similar to the first machine learning task, it is only the frustration model which includes feedback as part of its features. It is probable that feedback was also not included in the list of features because there was more relation between other features and the changes in excitement values. The better performance of linear regression over the other algorithms similar to what was stated in the previous machine learning task, may probably be due to its weighting capability and the inability of k-NN and SVM to work well given the nature of the data.

$$- 0.004 * \mathit{Gender} - .001 * \mathit{Feedback} + 0.826 * \mathit{Frustration} - 0.825 * \mathit{Excitement} - 0.473 * \mathit{SDEx2} + 0.339 * \mathit{SDFr10} + 0.003 * \mathit{Extroversion} - 0.004 * \mathit{Inquisitiveness} + .003 * \mathit{Accommodation} + 0.185 * \mathit{AveEx2} + 0.195 \tag{9.3}$$

$$0.909 * \mathit{Excitement} - 1.080 * \mathit{Frustration} + 0.003 * \mathit{Extroversion} + 0.417 * \mathit{AveFR2} + 0.264 * \mathit{SDFR5} + 0.002 * \mathit{Inquisitiveness} + 0.019 * \mathit{Age} + 0.222 * \mathit{SDEx5} + 0.184 * \mathit{SDFr10} + 0.220 * \mathit{SDEx2} - 0.748 \tag{9.4}$$

The models created by this machine learning task allow the prediction of the change in frustration and excitement after feedback is given. Change can either be an increase or decrease in frustration and excitement as well as its magnitude. This now allows feedback impact to be measured automatically and can be used for selecting appropriate feedback. This will enable future ITS to identify the impact of feedback, then revise those that seem suboptimal and retain other feedback that are optimal.

10. Conclusion and Future Work

The frustration and excitement models created in this research show the potential of using emotions extracted from brainwave signals for predicting a student's emotions resulting from the appraisal of feedback in an ITS. Other researches identify emotions at a specific time; however, the predictive models created in this research allow emotions to be attributed to feedback. The use of these predictive models will allow future ITS to predict what a student's emotion will be after appraising feedback. In the case that it is not favorable, the feedback can be modified or retained if otherwise.

By identifying emotions extracted from brainwave signals and using system logs, there is no need to interrupt the student's learning task to request self-reported affective evaluations of the feedback given. It also removes the need for expert annotation of student's emotions removing biases and human error.

We are currently working on improving the feature set used for predicting feedback appraisal. We feel that there may be better features that can be used to improve accuracy such as descriptions of the current context and history of activities in the ITS. By providing more descriptive features we hope to increase the performance of the existing predictive models. It will also be interesting to investigate feedback appraisal on individual students instead of using a more general approach. Lastly, we are working on integrating this technology with an existing ITS to allow automatic feedback adjustment using both cognitive states and predicted feedback appraisal.

Acknowledgments

This work is supported in part by the Philippine Council for Advanced Science and Technology Research - Department of Science and Technology through the Center for Empathic Human-Computer Interactions of De La Salle University, Manila, Philippines; by the Management Expenses Grants for National Universities Corporations through the Ministry of Education, Culture, Sports, Science and Technology of Japan; and by the Japan Society for the Promotion of Science.

We would like to acknowledge the students who participated in the experiments and Mr. Gian Fontanilla for helping us facilitate the experiment with his classes. We would also like to thank Dr. Geoffrey Mackellar of Emotiv for sharing with us information regarding the Emotiv EPOC Neuroheadset and inputs regarding our research.

References

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. In *Proc. 2009 Conference on Artificial Intelligence in Education* (pp. 17–24). Amsterdam, The Netherlands: IOS Press.

Arroyo, I., Muldner, K., Burleson, W., Woolf, B., & Cooper, D. (2009). Designing affective support to foster learning, motivation and attribution. In *Proc. Closing the Affective Loop in Intelligent Learning Environments Workshop at the 14th International Conference on Artificial Intelligence in Education*. Brighton, UK.

Aurellano, M., Tabuñar, B., Tan, R., Tuason, E., & Suarez, M. (2007). POOLE III: Using a Bayesian network for individual and group student modelling. In *Proc. 7th Philippine Computing Science Congress*. Boracay, Philippines.

Burleson, W., & Picard, R. W. (2004). Affective agents: Sustaining motivation to learn through failure and a state of “stuck”. In *Proc. Social and Emotional Intelligence in Learning Environments Workshop in conjunction with the 7th International Conference on Intelligent Tutoring Systems*. Maceio - Alagoas, Brasil.

Chan, E., Pulido, O., Torchiva, L., Uy, R., & Cruz, M. (2004). Programmers’ object-oriented learning environment. In *Proc. 4th Philippine Computing Science Congress*. Laguna, Philippines.

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, *19*(3), 267–303.

D’Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., Kaliouby, E. R., Picard, R. W., & Graesser, A. (2008). Autotutor detects and responds to learners affective and cognitive States. In *Proc. Workshop on Emotional and Cognitive Issues at the International Conference on Intelligent Tutoring Systems*. Montreal, Canada.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *The American Psychologist*, *48*(1), 26–34.

Graesser, A. C., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S., & Gholson, B. (2006). Detection of emotions during learning with autotutor. In *Proc. 28th Annual Meetings of the Cognitive Science* Society (pp. 285–290).

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley and Sons.

Kapoor, A., & Picard, R. W. (2005). Multimodal affect recognition in learning environments. In *Proc. 13th Annual ACM International Conference on Multimedia (MULTIMEDIA'05)* (pp. 677–682). New York, NY: ACM.

Legaspi, R., Fukui, K., Moriyama, K., Kurihara, S., Numao, M., & Suarez, M. (2010). Addressing the problems of data-centric physiology-affect relations modeling. In *Proc. 14th International Conference on Intelligent User Interfaces* (pp. 21–31). New York, NY: ACM.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). YALE: Rapid prototyping for complex data mining tasks. In L. Ungar, M. Craven, D. Gunopulos, & T. Eliassi-Rad (Eds.), *Proc. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)* (pp. 935–940). New York, NY: ACM.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D., & Strohecker, C. (2004). Affective learning - A manifesto. *BT Technology Journal*, *22*(4), 253–269.

Pour, P. A., Hussain, M., Alzoubi, O., D'Mello, S., & Calvo, R. (2010). The impact of system feedback on learners' affective and physiological states. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent Tutoring Systems* (Lecture Notes in Computer Science , Vol. 6094, pp. 264–273). Berlin / Heidelberg: Springer.

Robison, J., Mcquiggan, S., & Lester, J. (2010). Developing empirically based student personality profiles for affective feedback models. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent Tutoring Systems* (Lecture Notes in Computer Science, Vol. 6094, pp. 285–295). Berlin / Heidelberg: Springer.

Scherer, K. R., Shorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion: Theory, methods, research*. Canary, NC: Oxford University Press.

Siemer, M., Mauss, I., & Gross, J. J. (2007). Same situation - different emotions: How appraisals shape our emotions. *Emotion*, *7*(3), 592–600.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199–222.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley-Interscience.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548–573.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (Second Edition). San Francisco, CA: Morgan Kaufmann Publishers.

Woolf, B. P., Beck, J., Eliot, C., & Stern, M. (2001). Growth and maturity of intelligent tutoring systems: A status report. In K. D. Forbus, & P. J. Feltovich (Eds.), *Smart machines in education* (pp. 99–144). Cambridge, MA: MIT Press.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognising and responding to student affect. *International Journal of Learning Technology*, *4*(3/4), 129–164.

Zakharov, K. (2007). *Affect recognition and support in intelligent tutoring systems*. Unpublished master thesis, University of Canterbury, New Zealand.