Inferring Students' Emotions Using a Hybrid Approach that Combine Cognitive and Physical Data

Ernani Gottardo[1(✉)] and Andrey Ricardo Pimentel[2]

[1] Federal Institute of Education, Science and Technology of Rio Grande do Sul, IFRS, Bento Gonçalves, Brazil
`ernani.gottardo@erechim.ifrs.edu.br`
[2] Federal University of Paraná, UFPR, Curitiba, Brazil
`andrey@inf.ufpr.br`

Abstract. There are nowadays a strong agreement in the research community that emotions directly impact learning. Then, an important feature for a software that claim to be useful for learning is to deal with students' affective reactions. These software should be able to adapt to the users' affective reactions, trying to get a more natural human-computer interaction. Some priors works achieved relative success in inferring students' emotion. However, most of them, rely on intrusive, expensive or little practical sensors that track students' physical reactions. This paper presents as its main contribution the proposal of a hybrid model for emotion inference, combining physical and cognitive elements, using cheaper and little intrusive method to collect data. First experiments with students in a traditional classroom demonstrated the feasibility of this proposal and also indicated promising results for inference of learning centered emotions. In these experiments we achieve an accuracy rate and Cohen Kappa near to 65% and 0.55, respectively, in the task of inferring five classes of learning centered emotion. Even though these results are better than some related work, we believe they can be improved in the future by incorporating new data to the model.

Keywords: Emotion inference · Learning related emotion · Affective tutoring · Affective computing

1 Introduction

Using computers as a tool to support educational process is not new and this subject has received attention from scientific community in recent years. Even so, many issues remain related to effectiveness and possible contributions of these environments to improve the learning process [1]. One of the main limitation of educational software available nowadays refers to the lack of features to customize or adapt the software according to individual needs of the learner [2].

Intelligent Tutoring Systems (ITS) try to overcome these limitations by implementing adaptive features based on learners' individual needs. However, one of the main gap presented by most of ITS available today is the absence of features to adapt to the emotional states of the students [1, 3].

Ignoring students' affective reactions is an important shortcoming of an educational software considering that cognition and neuroscience researches largely agrees that emotions play a fundamental role in humans behavior [4]. Emotions directly influence simple and automatic activities, such as reaction to some threatening event, to more complex activities such as decision making and learning [3].

According to [5], good teachers are experts in observing and recognizing students' emotional states and, based on this recognition, take actions trying to positively impact learning. Furthermore, studies indicate that approximately half of the interactions between human tutors and apprentices has focused on aspects related to affective and engagement issues [6]. Thus, as observed by [4], computational environments unable to recognize emotions are severely restricted, especially in tasks such as learning or tutoring.

So, investigating the impact on learning of factors such as motivation and affective states has emerged as a promising research avenue. Previous works have shown that usability improvements obtained by computing environments that are able to infer and adapt to affective students' reactions [7].

As an example, an educational software should not interrupt learners who are progressing well, but could offer help to others who demonstrate a steady increase in frustration [8]. Another example is to implement pedagogical intervention strategies, seeking to avoid the so-called vicious cycle [3]. Vicious cycle is characterized by the repetition of affective states with negative valence that make learning difficult, such as boredom and frustration.

Of course, the computing environment can allow the user to make on-demand requests to adapt to their needs. However, it has been shown that better results regarding the usability of the system [7] or better learning experience [9] are reported with the use of systems that proactively adapt to their users. In this context, to provide any kind of adaptation to users' emotions, it's first necessary that emotions are properly recognized by the computational environment. The task of automatically inferring users' emotions by computers is a hard job and still presents several barriers and challenges to overcome [10, 11]. The challenges range from conceptual definitions related to emotions, to mapping of signals and computationally treatable patterns into emotions [4]. In order to overcome these challenges, researchers have used a wide range of techniques and methods from a relatively new research area, known as Affective Computing.

In this context, this paper presents the proposal of a hybrid model for inference of students' emotion while using a computing learning environment. This model allows us to investigate how quite distinct data modalities (eg physical reactions and contextual information) can be combined or complemented each other. The main goal of this proposal is to improve the emotion inference process, trying to fill an important gap in current research in which the proposed hybrid approach is little explored. It is also worth noting that low-cost and non-invasive strategies (logs of system events) and minimally invasive (facial expressions) strategies are used to obtain data for the model.

The results presented in this paper point to the technical feasibility of the proposed model. In addition, some promising results could be obtained in the inference process. These results illustrate how the physical and cognitive component of the model interact in the classification task for five classes of learning related emotions considered in this work.

2   Conceptual Bases and Correlated Works

The Hybrid Model of Emotions Inference - ModHEmo proposed in this work fits in a research area called 'Affective Computing'. Affective Computing is a multidisciplinary field that uses definitions related to emotions coming from the areas like psychology and neuroscience, as well as computer techniques such as artificial intelligence and machine learning [4].

Beyond the application in educational environments, which is the focus of this research, affective computing techniques have been used in areas such as entertainment, marketing, medicine, games, human-computer interaction, among others.

The proposal presented in this work relates with one of the areas of affective computing that deals with the challenge of recognizing humans' emotions by computers. However, it is important to emphasize that human emotions or affective states are not directly observable or accessed [4]. What is revealed voluntarily or involuntarily by people are patterns of expressions and behaviors. Considering these patterns, people or systems can apply computational techniques to infer or estimate the emotional state, always considering a certain level of error or uncertainty.

Building computing environments able to recognize humans emotions has proved to be a challenge. The main obstacle is the high level of ambiguity in the process of mapping between affective states and the signal data that can be used to detect them. In this sense, it should be noted that human beings also present difficulties and ambiguities in the recognition of other people's emotions [12].

Some assumptions presented in [4] were used as a base for the construction of the ModHEmo. This author advocates that an effective process of emotions inference should take into account three steps or procedures that are common when a person tries to recognize someone else's emotions. These three steps are: (I) identify low-level signals that carry information (facial expressions, voice, gestures, etc.), (II) detect signal patterns that can be combined to provide more reliable recognition (e.g., speech pattern, movements) and (III) search for environmental information that underlies high level or cognitive reasoning related to what kind of emotional reaction is common in similar situations.

Considering the three steps or procedures described above and the correlated works consulted, we observed that several studies have been based only on the step I or I and II. Much of this research makes the inference of emotions based on physiological response patterns that could be correlated with emotions. Physiological reactions are captured using sensors or devices that measure specific physical signals, such as the facial expressions (used in this work). Among these devices, it may be mentioned: sensors that measure body movements, [10], heartbeat [4,10], gesture and facial expressions [2,10,13,14], skin conductivity and temperature [4].

On the other hand, some research like [12,15,16] use a cognitive approach, heavily relying on step III, described above. These researches emphasize the importance of considering the cognitive or contextual aspects involved in the process of generation and control of humans' emotions. In this line, it is assumed that the emotions are activated based on individual perceptions of positive or negative aspects of an event or object.

The relevance of considering cognitive/contextual aspects together with physical reactions is illustrated by the following three examples: (I) tears can be recognized

from a video of the face, however it does not necessarily correspond to sadness, and may also represent joy [4], (II) emotions with negative valence tend to increase heart rate, but heart rate alone provide little information about specific emotions, and (III) research shows that affective states such as frustration or annoyance are not clearly distinguishable from neutral affective state using only facial expressions [17, 18].

As we can be seen below in Sect. 5, the hybrid model proposed in this work stands out by simultaneously integrating physical and cognitive elements, which are naturally integrated by humans when inferring someone else's emotions.

3   Adapting the Computational Environment Based on Affective Inferences

It is important to note that recognizing emotions represents only the first step toward creating computational environments that are adaptable to the affective reactions of its users. However, as pointed out by [3], the correct identification of affective states is indispensable for the development of *affect-sensitive* computing environments.

Considering the educational domain, the work of [19] shows that approximately half of the interactions between human tutors and apprentices focus on aspects related to affective and engagement issues. For example, a good teacher, realizing that students are confused or frustrated, should revise their teaching strategies in order to meet learners' needs, which could lead to improvements in learning [2]. On the other hand, offering help that disrupts the concentration or engagement of a student may be harmful [4].

In order to simulate the behavior of a good human tutor, improvements in the process of interaction and usability of educational software could be implemented. An alternative to this is the adaptation of the environment based on tutorial intervention strategies that make use of information about the affective states of the students.

An example of the importance of adaptation is presented in [20] which note that when the affective state confusion is not properly monitored and managed the student may become bored, affective state that hampers or even impedes learning.

In this direction, [8] presents a tutorial intervention strategy that combines cognitive and affective elements. Table 1 presents some examples of this strategy, containing the cognitive and affective element that guide the choose of the most appropriate intervention strategy.

Mentoring interventions could also be used to avoid what [21] call a 'vicious cycle'. The 'vicious cycle' occurs when one or more negative cognitive-affective states recur repeatedly, indicating that interventions are necessary to assist the students in order to help them overcome potential difficulties.

4   Set of Emotion to Be Inferred

Related work [3, 20, 22, 23] has shown that some emotions have a greater impact on the learning process. However, does not exist yet a complete understanding of which emotions are the most important in the educational context and how they influence learning [22, 23].

Table 1. Adaptive strategies combining cognitive and affective elements [8].

Cognitive element	Affective element	Adequate tutorial intervention
Student makes a mistake	Student demonstrates be curious and focused	No intervention is required. Student is engaged in learning and exploration *flow*)
	Student frowns, becomes restless and look around	Intervention actions are required. Student is confused
Student is solving problems correctly	Student shows no signs of frustration or annoyance	No intervention is required. Student is in control, concentrated and focused
	Student shows boredom or disappointment (very easy problems)	Intervention is necessary, possibly increasing the difficulty of the problems

Even so, affective states such as confusion, annoyance, frustration, curiosity, interest, surprise, and joy have emerged in the scientific community as highly relevant because their direct impact in learning experiences [23].

Considering that this research focus on the application in an educational scenario, firstly it was evaluated which set of emotions should be considered in the inference process. In this context, choosing the set of emotions to be included in this work was carried out seeking to reflect relevant situations for learning. Thus, the 'circumplex model' of [24] and the 'spiral learning model' of [5] were used as reference. These theories have been consolidated and frequently referenced in related works such as [3, 12, 23, 25].

To choose the set of emotion, we also considered a mapping of learning-centered cognitive–affective states into a two integrated dimensions: valence (pleasure to displeasure) and arousal (activation to deactivation) as shown in Fig. 1. This mapping is presented in [3] and is based on Russell's Core Affect framework [26].

Taking into account the arguments presented so far in this section, it was understood that a rational, efficient and innovative approach could be making inferences based not in a specific set of emotions, but grouping correlated set of emotions into quadrants.

In this way, the Fig. 2 shows the approach used in this work to arrange the emotions related to learning. In this proposal the dimensions 'Valence' (positive or negative) and 'Activation' or intensity (agitated or calm) are used for representing emotions in quadrants named as: Q1, Q2, Q3 and Q4. It was also assigned a representative name (see Fig. 2) for each of the quadrants considering learning related states. To represent the neutral state its was create a category named QN, denoting situations in which both valence and activation dimensions are zeroed. These quadrants, plus the neutral stated, played the role of classes in the classification processes performed by the ModHEmo that will be described in the next section.

The proposal shown in Fig. 2 is aligned with the assumptions made by [5] that teachers adapt themselves to assist students based on a small set of affective reactions as opposed to a large number of complex factors. These authors suggest that it would be advisable to work with a set of simplified emotions and that this set could be refined considering the advances in the research.
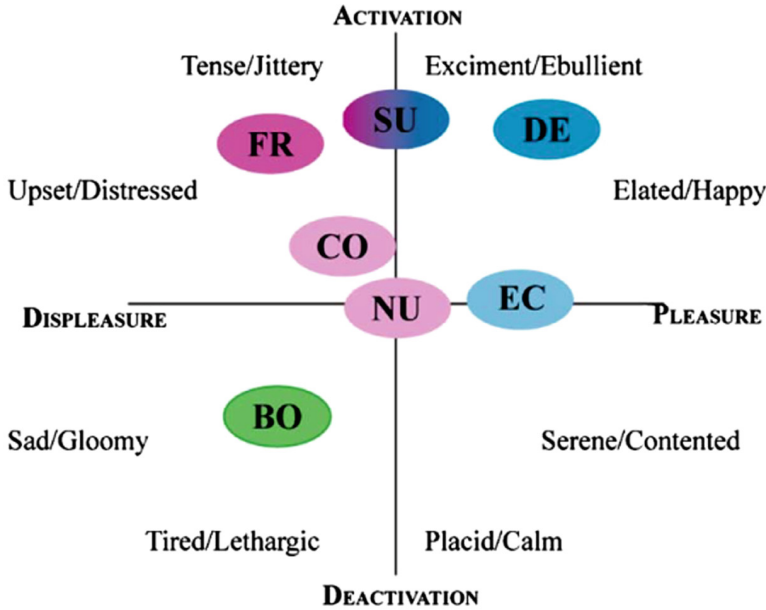
BO: Boredom, EC: Engaged Concentration, FR: Frustration, DE: Delight, SU: Surprise, NU:Neutral

Fig. 1. Learning-centered cognitive–affective states [3].

Although not considering individual emotions, it is understood that the identification of the quadrants, plus the neutral state, can be a very important aspect and enough to support many kind of adaptation actions and intervention strategies, for example.

Figure 2 also shows the main single emotions contained in each quadrant, divided into two groups: (i) physical and (ii) cognitive. These groups represent the two distinct type of data sources considered in ModHEmo. Each emotion was allocated in the quadrants considering its values of the valence and activation dimensions. The values for these two dimensions for each emotion were obtained in the work of [28] for the cognitive emotions and in [25] for the physical emotions.

The emotions in the physical group are the eight basic or primary emotions described in the classic model of [29] which are: anger, disgust, fear, happy, sadness, surprise, contempt and neutral. This set of emotions are inferred through the students' facial expressions observed during the use of an educational software.

The emotions in the cognitive group are based on the well-known cognitive model of Ortony, Clore e Collins - OCC [30]. The OCC model is based on the cognitivist theory, explaining the origins of emotions and grouping them according to the cognitive process that generates them. The OCC model consists of 22 emotions. However, based on the scope of this work, eight emotions were considered relevant: joy, distress, disappointment, relief, hope, fear, satisfaction and fears confirmed. These set of emotions were chosen because, according to the OCC model, its include all the emotions that are triggered as a reaction to events. The kind of events that occurs in the interface of an educational software are used to infer the cognitive emotions.
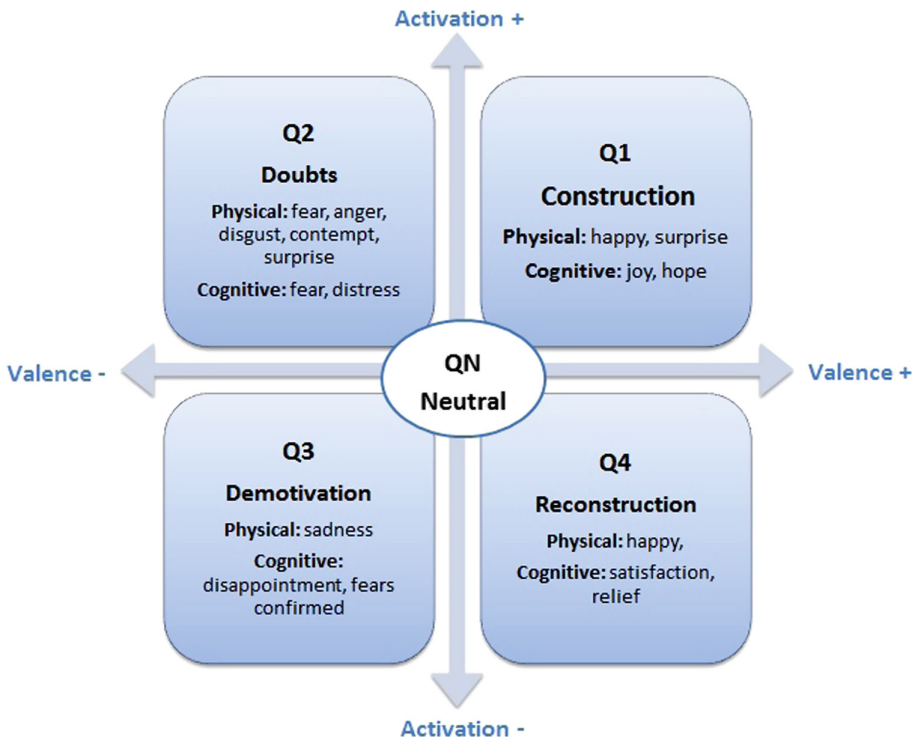
Fig. 2. Quadrants and learning related emotions [27].

It can be observed in Fig. 2 that the "happy" and "surprise" emotions in the physical component appear repeated in two quadrants, as they may have high variability in the activation and valence dimension, respectively. To deal with this ambiguity, in the implementation of the hybrid model described below, its observed the intensity of the happy emotion inferred: if happy has a score greater than 0.5 it was classified in the Q1 quadrant and, otherwise, in the Q4 quadrant. For 'surprise' emotion, which may have positive or negative valence, the solution used in the implementation of the model was to check the type of event occurring in the computational environment: if the valence of the event is positive (e.g. correct answer) 'surprise' was classified in the quadrant Q1 and, otherwise, in the Q2 quadrant.

Furthermore, the OCC model does not include a neutral state. So to infer this affective state we considered the condition when the scores of all the eight emotions of the cognitive component are equal to zero.

5   Hybrid Model of Emotion Inference

In order to evaluate the feasibility and the performance of this proposal, a hybrid model for inference of affective states - ModHEmo was proposed and implemented. This model will serve as a basis for structuring the research development and also for evaluating its results. Figure 3 schematically shows the proposed model. The main feature of
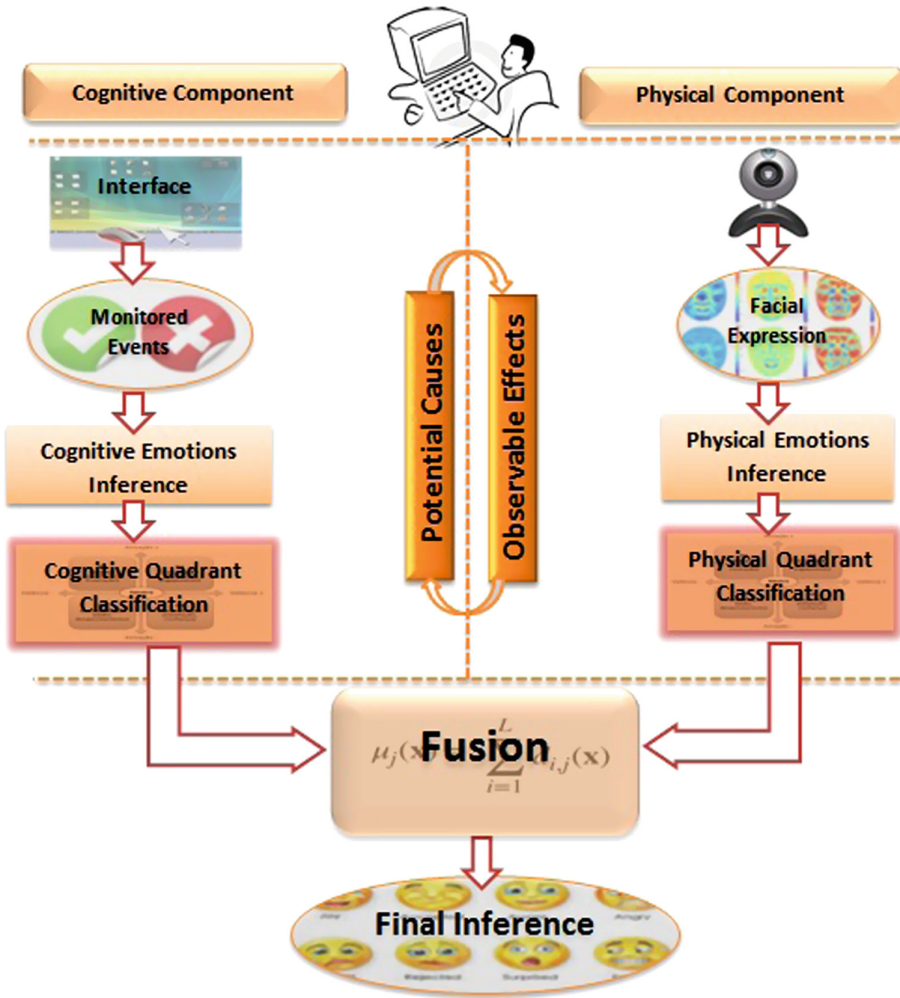
Fig. 3. Hybrid model of learning related emotions inference [27].

ModHEmo is the initial division of the inference process into two fundamentals components: physical and cognitive. This figure also shows the modules of each component and the fusion of the two components to obtain the final result.

The cognitive component, based on the OCC theory, is responsible for managing the relevant events in the computing environment. In order to implement cognitive inference, a custom version of the ALMA (A Layered Model of Affect) [28] model was used. The process of cognitive emotions inference returns *scores* normalized in the interval [0,1] for each of the eight cognitive emotions.

The physical component of ModHEmo deals with observable reactions using student's face images. Images are obtained using a standard webcam, following the occurrence of a relevant event in the interface of the computing environment. These images are used to infer the eight physical emotions included in ModHEmo by using the

EmotionAPI tool[1]. This inference process also returns *scores* normalized in the interval [0,1] for each of the eight emotions of the physical component.

Based on these initial inferences, a classification process is performed to map the emotions to the quadrants depicted in Fig. 2. At the end of this step, a normalized score in the interval [0,1] is obtained for each quadrants and also for the neutral stated in the two components.

The *Softmax* function [31] showed in Eq. 1 was the method used to normalize in interval [0,1] the ModHEmo's cognitive and physical score results. In this equation, $g_1(x), ..., g_c(x)$ are the values returned for the eight emotion for each ModHEmo's components. Then, in Eq. 1 new scores values are calculated by $g'_1(x), ..., g'_c(x), g'_j(x) \in [0, 1], \sum_{j=1}^{C} g'_j(x) = 1$.

$$g'_j(x) = \frac{exp\{g_j(x)\}}{\sum_{K=1}^{C} exp\{g_k(x)\}} \tag{1}$$

For the fusion of the two ModHEmo components it is assumed that some combination technique is applied. For simplicity reasons, in a first version of ModHEmo the fusion process was implemented using the *sum* function. Thus, if we denote the scores assigned to class i by the classifier j as s_i^j, then a typical combining rule is a f function whose combined final result for class i is $S_i = f(\{s_i^j j = 1, ...M)$. The final result is expressed as $argmax_i\{S_i, ..., S_n\}$ [32]. In the context of this work, j plays the role of the physical and cognitive components while i is represented by the five classes depicted in Fig. 2.

Trying to improve the ModHEmo performance, a second version of ModHEmo was implemented. In this version, the final fusion process is performed by creating a single data set containing the scores of the quadrants of each component. After merging the two components, this dataset contains 10 attributes (5 physical + 5 cognitive) with quadrant scores for each component. The class obtained through the labeling process (to be described in the next section) is also part of the data set. Based on this data set, we trained RandomForest [33] e IBK [33] classification algorithms to perform the final inference of the model.

These algorithms were chosen because results in some initial tests with the database used in this work indicated that classification algorithms based on simple and fast techniques such as decision trees (RandomForest) and K-nearest neighbors (KNN) reached the best results.

As the main goal of this work doesn't include tuning of algorithms' parameters, we used the default Weka[2] parameters values. In addition, the 10-fold Cross-validation was used for splitting the base into test and training, considering that it's a robust and highly used technique [33].

In the next section, we will present detailed results obtained in experiments using the two versions of ModHEmo described above.

---

[1] https://azure.microsoft.com/services/cognitive-services/emotion/ developed by the University of Oxford and Microsoft.

[2] https://www.cs.waikato.ac.nz/ml/weka.

6   Experiment and Results

To verify the performance of ModHEmo inferences, experiments were performed using the two version of the model. In this experiments we used 'Tux, of Math Command' or TuxMath[3], an open source arcade game educational software. TuxMath is an educational game that allows kids to exercise their mathematical reasoning. In this game, the challenge is to answer math equation (four basic operations) to destroy meteors and protect the igloos and penguins.

The level of the game was chosen considering the age and math skill of the students. Within each level, the comets in TuxMath are released in waves with an increasing number of comets (2,4,6,8,10 ..) in each wave. A new wave begins only after all the comets of the previous wave are destroyed or reach the igloos.

Its important to notice that the experiments were approved and follows the procedures recommended by the ethics committee in research of the public federal educational institution in which the first author is professor.

Trying to reduce the interference of the research in students' behavior, the experiment was carried out in the computer lab normally used by the students and with the presence of their teacher. In addition, the students were instructed to perform the activity in a natural way without restrictions of position, movements, etc.

While students used TuxMath, some of the main events of the game were monitored. Among these events, it can be highlighted: correct and wrong math equation answers, comets that damaged penguins' igloos or killed the penguins, game over, win the game, etc.

Additionally, in order to artificially create some situations that could generate emotions, a random bug generator procedures was developed in Tuxmath. Whenever a bug was artificially inserted, it would also become a monitored event. These bugs include, among others, situations such as: (i)non-detonation of a comet even with the correct math equation response, (ii) display of comets in the middle of the screen, decreasing the time for the student to enter the correct answer until the comet hits the igloo or penguin at the bottom of the screen. The students were only informed about these random bugs after the end of the game.

These events was used as input to the cognitive component of the ModHEmo. Following the occurrence of a monitored event in TuxMath, student's face image was captured with a basic webcam and this image is used as input to the physical component of ModHEmo. With these inputs the model is then executed, resulting in the inference of the probable affective state of the student in that moment.

Before starting the game, two questions are presented to student to gather information about their goals and prospects. The first question ask students about their goals in the activity. Only data of the 15 students with 'win the game/learn math'[4] goal was considered. The second question ask students about their prospect: win or lost the game. With this information it was possible to set ALMA Tags like 'GOODEVENT', 'BADEVENT', etc.

---

[3] http://tux4kids.alioth.debian.org/tuxmath/index.php.

[4] Other goals included: have fun, participate in research and compete with colleagues.

After completing the session in the game, students were presented with a tool developed to label the data collected during the experiment. This tool allows students to review the game session through a video that synchronously shows the student's face along with the game screen. The video is automatically paused by the labeling tool at the specific time that a monitored event has happened. At this moment an image with five representative emoticons (one for each quadrant and one for neutral state) is shown and asked the student to choose the emoticon that best represents their feeling at that moment. After the student's response, the process continues.

The Fig. 4 shows a screen of the tool developed to the labeling process described above. It is highlighted in this figure four main parts: (I) the upper part shows the student's face at timestamp 2017-10-02 12:36:42.450. (II) at the bottom it is observed the screen of the game synchronized with the upper part, (III) emoticons and main emotions representative of each of the quadrants plus the neutral state and (IV) description of the event occurred in that specific timestamp (Bug - Comet Displayed in the Middle Screen).

Affective assessment using emoticons was adopted based on the fact that children perceive the classical Self-Assessment Manikin (SAM) [34] approach difficult to use and understand, according to the results presented in [35,36]. Then, we choose to use emoticons that are familiar to kids because of its widespread use in social networks, message apps, etc.

In order to facilitate understanding about ModHEmo's operation, the results of a specific student using the first version of ModHEmo will be detailed below. The student id 6 (see Table 2) was chosen for this detailing because he was the participant with the highest number of events in the game session. Figure 5 shows the results of the inferences made by ModHEmo with student id 6. It is important to emphasize that in the initial part of the game depicted in Fig. 5 the student showed good performance, correctly answering the arithmetic operations and destroying the comets. However, in the middle part of the game, several comets destroyed the igloos and killed some penguins. But, at the end, the student recovered after capturing a power up comet[5] and won the game.

The lines in Fig. 5 depict the inferences of the physical and cognitive component and also the fusion of both. The horizontal axis of the graph shows the time and the vertical axis the quadrants plus the neutral state (see Fig. 2). The order of the quadrants in the graph was organized so that the most positive quadrant/class Q1 (positive valence and activation) is placed on the top and the most negative Q3 (negative valence and activation) on the bottom with neutral state in the center.

Aiming to provide additional details of ModHEmo's inference process, two tables with the scores of each quadrant (plus neutral state) in the physical and cognitive components were added to the Fig. 5 chart. These table show the values at the instant '10/05/2017 13:49:13' when a comet destroyed an igloo. It can be observed in the tables that in that instant, for the cognitive component, the quadrant with the highest score (0.74) was Q3 (demotivation), reflecting the bad event that has occurred (comet destroyed an igloo). In the physical component the highest score (0.98) was obtained

---

[5] Special kind of comet that enable a gun able to simultaneously destroy all the penguins in the screen.

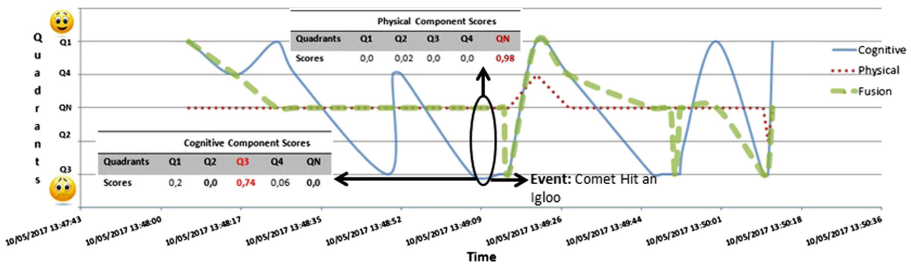**Fig. 4.** Screenshoot of labeling tool [27].

Fig. 5. ModHEmo results in one game session [27].

by QN state (neutral) indicating that student remained neutral, regardless of the bad event. Considering the scores of these tables, the fusion process is then performed. For this, initially the scores of the physical and cognitive components for each quadrant and neutral state are summed and the class that obtained greater sum of scores is chosen. As can be seen in the graph, the fusion process at these instant results in neutral state (QN), which obtained the largest sum of scores (0.98).

In the game session shown in Fig. 5 it can be seen that the physical component of the model has relatively low variation remaining most of the time in the neutral state. On the other hand, the line of the cognitive component shows a greater amplitude including points in all the quadrants of the model. The fusion line of the components remained for a long time in the neutral state, indicating a tendency of this student to not negatively react to the bad events of the game. However, in a few moments, the fusion line presented some variations accompanying the cognitive component.

6.1   Results Using *Sum* Function as a Fusion Technique

In an experiment with the first version of ModHEmo using the *Sum* function as fusion technique, eight elementary students with age ranging from ten to fourteen years old played the game.

Using the data collected with the labeling process describe above, it was possible to check the accuracy of the inferences made by ModHEmo. For example, considering the student id 6, the Fig. 6 shows the accuracy of ModHEmo inferences. This Figure shows two lines depicting the fitness between values of labels and ModHEmo inferences using the data of student id 6. For this student, the accuracy rate was 69%. So, the inferences were correct in 18 of 26 events for this students' playing session.
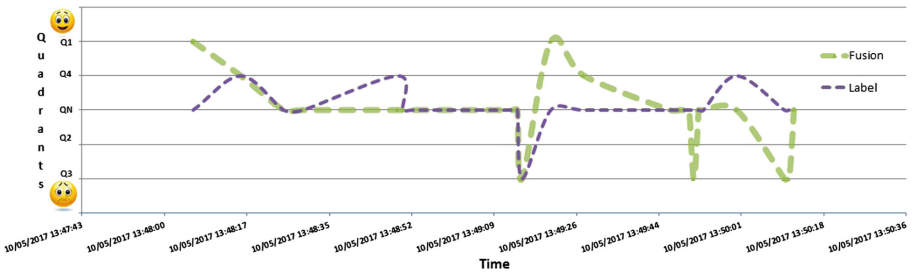

Fig. 6. Comparision between ModHEmo inferences and labels [27].

Table 2 shows the results of the eight students participating in the experiment. This table shows the number of monitored events, the number of correct inferences of the ModHEmo (Hits) and the percentage of accuracy. The number of monitored events in Table 2 is variable due to the fact that it depends, among others, on the game difficulty and student performance in the game. For confidentiality reasons, Table 2 shows only a number as students' identification. Student 6 data was used in the examples of Figs. 5 and 6 above.

6.2   Results Using Classifiers Algorithms as Fusion Technique

Trying to improve the accuracy of the ModHEmo inferences, a second version of the model was developed using classification algorithms as a fusion strategy of the physical and cognitive components.

Table 2. ModHEmo prediction accuracy [27].

Student	#Events	#Hits	% Accuracy
1	15	7	47
2	18	10	56
3	9	5	56
4	9	8	89
5	14	7	50
6	26	18	69
7	11	6	55
8	8	5	63
Total	110	66	60

To test this new version a second experiment was conducted. In this experiment participated a total of 15 students, with ages ranging from 10 to 14 years. These students were enrolled between the fifth and ninth year of elementary school in the Jaguaretê Municipal School of Education, in the rural area of Erechim-RS, Brazil.
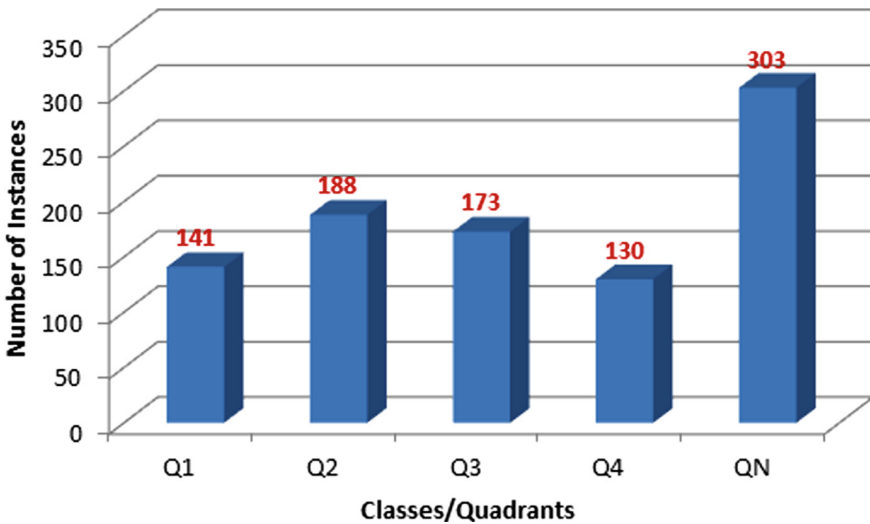

Fig. 7. Distribution of the ground truth dataset.

It is important to note that among the 15 students participating in this second experiment, it was included the data of the 8 students participating in the first experiment. This approach was used in order to achieve a larger dataset that should be more suitable for the task of training and testing the classification algorithms.

In this experiment we gather a dataset with 935 instances of monitored events in Tuxmath. Using the labeling tool described in the previous section, these 935 events were labeled by the 15 students. In this way we get a ground truth dataset that was used to train and test the classification algorithms. The Fig. 7 shows a chart with a distribution of the of the ground truth dataset created in the experiment with 935 instances.

As we can see in Fig. 7 the Class QN (neutral) was the most frequent with 303 instances (32.4%). On the other side, the class Q4 (reconstruction) was the less frequent with 130 instances (13.9%).

Table 3 shows some results achieved in the experiment using the labeled dataset with 935 instances. In this table, it's presented frequently used performance classifier metrics [33] for each algorithms.

Table 3. Global performance metrics of the two classifiers.

Algorithm	Accuracy (%)	Accuracy (instances)	Cohen's Kappa	F-Measure
RandomForest	64.81	606	0.545	0.648
IBK	63.53	594	0.532	0.635

Table 3 show that RandomForest algorithm achieved a global accuracy of 64.81% while IBK accuracy was 63.52%. The Pair-Wise T-Test [33] with significance level of 0.05 executed in Weka Experiment Environment showed that there is no statistical significant difference in accuracy between these two algorithms.

In the sequence, the Table 4 presents the confusion matrix [33] with the results of the RandomForest algorithm. The main diagonal of the confusion matrix, highlighted with gray background, represents the correct inferences for each of the five classes considered in this experiment.

Table 4. Confusion matrix for RandomForest algorithm.

Classe Correta \ Classe Prevista	Q4	Q2	Q3	QN	Q1
Q4	70	10	10	28	12
Q2	4	131	25	27	1
Q3	13	24	95	37	4
QN	13	27	37	216	10
Q1	15	12	4	16	94

The ROC (Receiver Operating Characteristic) curves [33] are a tool frequently used to check the performance of a classifier without regard to class distribution or error costs.

The Fig. 8 depict the ROC curves for the five classes obtained by the RandomForest algorithm (the curves for IBK are very similar). Inside this figure we also add a table that shows the Area Under Curve (AUC) computed in Weka. AUC report accuracy as and index ranging from zero to one (1 for a perfect and $<= 0.5$ a random classifier) [33].

6.3 Analysis of the Experiments Results

Analyzing the results of the two experiments reported above, it can be observed that in terms of global accuracy there was no significant change between the experiments. This could be considered an indication of generalization of the proposed model. However, experiments with more students are necessary to confirm the generalization of the ModHEmo results.

The task of comparing the results presented in this paper with correlated works is a sensitive and hard task because a lot of factors, like: (i) the types of sensors used in the experiments, (ii) experiment applied in a real environment or laboratory, (iii) type of interaction with the computing environment (text, voice, reading, etc.), (iv) who does the data labeling (students, external observers, etc.).
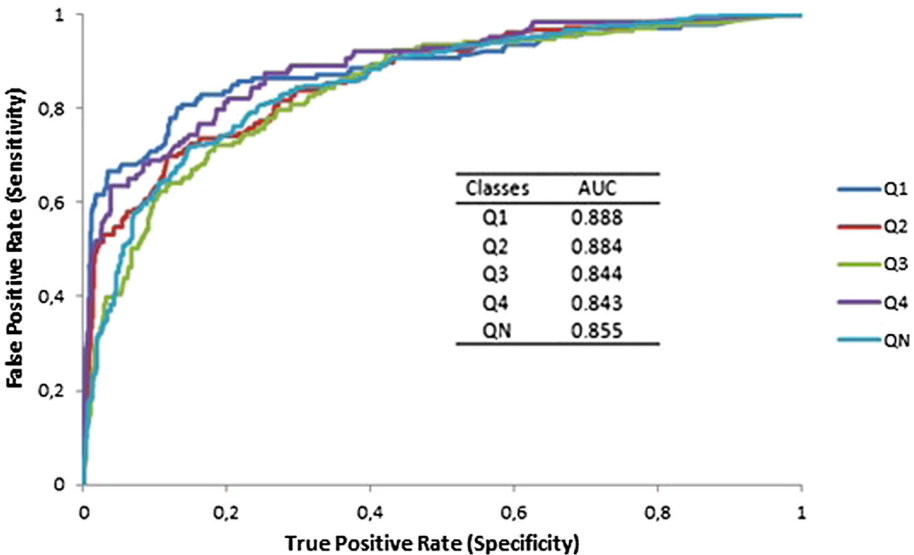

Classes	AUC
Q1	0.888
Q2	0.884
Q3	0.844
Q4	0.843
QN	0.855

Fig. 8. ROC curves and AUC for RandomForest algorithm [37].

For example, relating to the labeling process, the work of [21] show different accuracy rates achieved when distinct actors do the labeling. These authors reported accuracy rate near to 62% when external judges or colleagues labeled the data. However, the accuracy dropped for near 52% when the student's themselves labeled the dataset.

To the best of our knowledge, there are no priors works that use exactly the same approach, kind of emotions, experimental configuration, etc. as presented in this work.

Even so, we consider important to make some comparison trying to position our results within the state of art. In this way, recent research of [8, 38–40] resemble with this work because they focus on learning related emotion and use a similar experimental design.

The works of [8] and [21] has some similarities with the present proposal because they use an identical set of emotions and experiments are made in a real learning environment. In [8], accuracy rates reported range between 80% and 89%. However, this higher accuracy is achieved by fusing a set of expensive or intrusive physical sensors, including: highly specialized camera (Kinect), chair with posture sensor, mouse with pressure sensor and skin conductivity sensor. [21] reports accuracy rates between 55% and 65% using text mining to predict emotion of students using an Intelligent Tutoring System.

Based on students' facial expression while using an ITS, [38] reported best Cohen's Kappa of 0.112. Using text mining techniques [39] showed a method to infer four emotions: bored, confused, frustrated and concentrated. For these emotions, the AUC reported was respectively 0.767, 0.777, 0.762, 0.738 and the best results for Cohen's Kappa was 0.486. The work of [40] used a deep learning approach, based on logs of students' interaction in an ITS, to predict four learning related emotion:confused, concentrating, bored, frustrated. In this work the best AUC and Cohen's Kapa was 0.78 and 0.24 respectively. Its important to note that in [40] the labeling was made by external observers.

So, the results achieved in the experiment with ModHEmo presented above show some improvements when compared with [21, 38–40]. In our experiment, Cohen's Kappa index was 0.545 and 0.532 (see Table 3) for RandomForest and IBK, respectively and AUC value was between 0.843 and 0.888 (see Fig. 8).

7   Final Consideration and Future Works

This work presented a hybrid inference model of learning related emotions that uses little or non intrusive sensors with potential for large scale and out-of-the-lab use. Inferences obtained with this model could be very useful for implementing learning environments able to appropriately recognize and adapt to learners' emotional reactions.

The model described in this paper stand out by presenting a method to combine quite distinct information (physical and cognitive) that is little explored in the research community nowadays. In this way, this proposal try to present improvements in the emotion inference process through the integration of these two important components involved in the generation and control of human emotions.

The initial results obtained can be considered promising, since, even if a direct comparison is difficult, the results obtained are similar or superior to the state of the art. However, as the hybrid approach resembles the natural process of emotions inference, it presents great opportunities for future improvements by adding new data or sensors.

Furthermore, this work is distinguished by focusing on a set of emotions relevant in teaching and learning contexts. Thus, the inferences of the proposed model, besides allowing automatic adaptations in the computational environment, could be used to depict a profile of students' affect dynamics that could drive individual pedagogical intervention.

For example, the affective states represented by the quadrants could be used to identify the so-called 'vicious-cycle' [21] which occurs when affective states related to poor learning succeed each other repeatedly. In the context of this work this 'vicious-cycle' could be detected in case of constant permanence or alternation in the quadrants Q2 and Q3. In these cases, pedagogical strategies to motivate the student should be applied.

Analyzing the results of the student id 6 presented in the previous section it can be verified that no 'vicious-cycle' could be detected nor repeated occurrences or permanence in the Q2 or Q3 quadrants. Therefore, specific actions of the educational software would not be necessary or advisable for the students used as example.

Cognitive-affective tutorial intervention strategies also could be based on the results of ModHEmo. These intervention strategies should not be applied to students who are interested or focused on the activity, even if some mistakes occurs. Furthermore, for students with constant signs of frustration or annoyance (quadrants Q2 and Q3) educational software could try strategies such as challenge or a game trying to alleviate the effects of these negative states.

As future work wed intend to expand the current dataset adding experimental data involving more students with other age groups and also other types of educational environments. It is also intended to evaluate the result of the adding new information in the physical and cognitive components. In the physical component could be aggregated information that can be obtained through the camera such as nods, blinks, etc. Information from the keyboard and mouse movements could be included in the cognitive component.

References

1. Khan, F.A., Graf, S., Weippl, E.R., Tjoa, A.M.: Identifying and incorporating affective states and learning styles in web-based learning management systems. IxD&A **9**, 85–103 (2010)
2. Alexander, S.T.V.: An affect-sensitive intelligent tutoring system with an animated pedagogical agent that adapts to student emotion like a human tutor. Ph.D. thesis, Massey University, Albany, New Zealand (2008)
3. Baker, R.S., D'Mello, S., Rodrigo, M., Graesser, A.: Better to be frustrated than bored: the incidence and persistence of affect during interactions with three different computer-based learning environments. Int. J. Hum.-Comput. Stud. **68**, 223–241 (2010)
4. Picard, R.W.: Affective Computing, vol. 252. MIT Press, Cambridge (1997)
5. Kort, B., Reilly, R., Picard, R.W.: An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. In: IEEE International Conference on Advanced Learning Technologies, Proceedings, pp. 43–46. IEEE (2001)
6. Lajoie, S.P., Derry, S.J.: Motivational techniques of expert human tutors: lessons for the design of computer-based tutors. In: Computers as Cognitive Tools, pp. 83–114. Routledge (2013)
7. Becker-Asano, C., Wachsmuth, I.: Affective computing with primary and secondary emotions in a virtual human. Auton. Agent. Multi-Agent Syst. **20**, 32 (2010)

8. Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., Picard, R.: Affect-aware tutors: recognising and responding to student affect. Int. J. Learn. Technol. **4**, 129–164 (2009)
9. Baker, R.S.J., et al.: Adapting to when students game an intelligent tutoring system. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 392–401. Springer, Heidelberg (2006). https://doi.org/10.1007/11774303_39
10. Grafsgaard, J.F., Wiggins, J.B., Boyer, K.E., Wiebe, E.N., Lester, J.C.: Embodied affect in tutorial dialogue: student gesture and posture. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 1–10. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_1
11. Baker, R.S., et al.: Sensor-free automated detection of affect in a cognitive tutor for algebra. In: Educational Data Mining 2012 (2012)
12. Conati, C.: Combining cognitive appraisal and sensors for affect detection in a framework for modeling user affect. In: Calvo, R., D'Mello, S. (eds.) New Perspectives on Affect and Learning Technologies. LSIS, vol. 3, pp. 71–84. Springer, New York (2011). https://doi.org/10.1007/978-1-4419-9625-1_6
13. Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., Bigdeli, A.: "How do you know that i don't understand?" a look at the future of intelligent tutoring systems. Comput. Hum. Behav. **24**, 1342–1363 (2008)
14. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learn. Instr. **22**, 145–157 (2012)
15. Jaques, P.A., Pesty, S., Vicari, R.: An animated pedagogical agent that interacts affectively with the student. In: AIED 2003, Shaping the Future of Learning Through Intelligent Technologies, pp. 428–430 (2003)
16. Paquette, L., et al.: Sensor-free or sensor-full: a comparison of data modalities in multi-channel affect detection. International Educational Data Mining Society (2016)
17. McDaniel, B., D'Mello, S., King, B., Chipman, P., Tapp, K., Graesser, A.: Facial features for affective state detection in learning environments. Proc. Cognit. Sci. Soc. **29**, 467–472 (2007)
18. Para revisão, O.: Omitido para revisão. In: Anais do XXXVI Congresso da Sociedade Brasileira de Computação, SBC, pp. 557–566 (2016)
19. Lepper, M.R., Woolverton, M., Mumme, D.L., Gurtner, J.L.: Motivational techniques of expert human tutors: lessons for the design of computer-based tutors. Comput. Cognit. Tools **1993**, 75–105 (1993)
20. Reis, H., Alvares, D., Jaques, P., Isotani, S.: Analysis of permanence time in emotional states: a case study using educational software. In: Nkambou, R., Azevedo, R., Vassileva, J. (eds.) ITS 2018. LNCS, vol. 10858, pp. 180–190. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91464-0_18
21. D'Mello, S., Picard, R.W., Graesser, A.: Toward an affect-sensitive autotutor. IEEE Intell. Syst. **22**, 53 (2007)
22. Picard, R.W., et al.: Affective learning-a manifesto. BT Technol. J. **22**, 253–269 (2004)
23. Aghaei Pour, P., Hussain, M.S., AlZoubi, O., D'Mello, S., Calvo, R.A.: The impact of system feedback on learners' affective and physiological states. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010. LNCS, vol. 6094, pp. 264–273. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13388-6_31
24. Russel, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**, 1161–1178 (1980)
25. Posner, J., Russell, J.A., Peterson, B.S.: The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev. Psychopathol. **17**, 715–734 (2005)
26. Russell, J.A.: Core affect and the psychological construction of emotion. Psychol. Rev. **110**, 145 (2003)

27. Gottardo, E., Pimentel, A.R.: Hybrid model of emotions inference- an approach based on fusion of physical and cognitive informations. In: Proceedings of the 20th International Conference on Enterprise Information Systems - ICEIS 2018: vol. 2, pp. 441–450. SciTePress (2018)

28. Gebhard, P.: ALMA: a layered model of affect. In: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 29–36. ACM (2005)

29. Ekman, P.: An argument for basic emotions. Cognit. Emotion **6**, 169–200 (1992)

30. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1990)

31. Kuncheva, L.I.: Combining Pattern Classifiers: Methods and Algorithms. Wiley, New York (2004)

32. Tulyakov, S., Jaeger, S., Govindaraju, V., Doermann, D.: Review of classifier combination methods. In: Marinai, S., Fujisawa, H. (eds.) Machine Learning in Document Analysis and Recognition. SCI, vol. 90, pp. 361–386. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-76280-5_14

33. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco (2016)

34. Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psychiatry **25**, 49–59 (1994)

35. Yusoff, Y.M., Ruthven, I., Landoni, M.: Measuring emotion: a new evaluation tool for very young chidren. In: Proceedings of the 4th International Conference on Computing and Informatics. In: ICOCI (2013)

36. Hayashi, E., Posada, J.E.G., Maike, V.R., Baranauskas, M.C.C.: Exploring new formats of the self-assessment manikin in the design with children. In: Proceedings of the 15th Brazilian Symposium on Human Factors in Computer Systems, p. 27. ACM (2016)

37. Gottardo, E., Ricardo Pimentel, A.: Improving inference of learning related emotion by combining cognitive and physical information. In: Nkambou, R., Azevedo, R., Vassileva, J. (eds.) ITS 2018. LNCS, vol. 10858, pp. 313–318. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91464-0_33

38. Bosch, N., Chen, Y., D'Mello, S.: It's written on your face: detecting affective states from facial expressions while learning computer programming. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) ITS 2014. LNCS, vol. 8474, pp. 39–44. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07221-0_5

39. Paquette, L., et al.: Sensor-free affect detection for a simulation-based science inquiry learning environment. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) ITS 2014. LNCS, vol. 8474, pp. 1–10. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07221-0_1

40. Botelho, A.F., Baker, R.S., Heffernan, N.T.: Improving sensor-free affect detection using deep learning. In: André, E., Baker, R., Hu, X., Rodrigo, M.M.T., du Boulay, B. (eds.) AIED 2017. LNCS (LNAI), vol. 10331, pp. 40–51. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61425-0_4