Affective Artificial Intelligence in Education: From Detection to Adaptation

Emmanuel G. BLANCHARD, Boris VOLFSON, Yuan-Jin HONG, Susanne P. LAJOIE
ATLAS Laboratory, McGill Faculty of Education, Montréal (QC), CANADA
emmanuel.blanchard@mcgill.ca, volfson@gmail.com, yuan-jin.hong@mail.mcgill.ca, susanne.lajoie@mcgill.ca

Abstract. This paper reviews and integrates research that would be necessary to develop an AIED system able to detect and then appropriately react to an affective state of a learner. It addresses the nature of affect, methods to automatically detect affect, as well as the interplay between affect and learning-related cognition, and affective strategies that promote quality learning.

Keywords. Affect detection, affect and learning, affect-related strategies.

Introduction

Researchers agree that affect influences learning [5, 11, 14, 20]. Picard [40] has stressed the importance of considering affect in artificial intelligence research, thus contributing to the development of affective awareness in Artificial Intelligence in Education (AIED). Hence, scholars have begun to design learning environments with affect management abilities whereby affective phenomena experienced by learners are identified and appropriate responses are presented through technology to enhance learning [3, 7, 8, 9, 10, 13, 25, 35]. However, despite encouraging findings, affective management in AIED remains limited for many reasons:

1. Even with numerous studies focusing on affect detection, deduction of the affective profile of learners in real time remains a complex task.
2. Cognitive psychology has long explored the effects of affect on learning; despite this, there is no agreed upon method to integrate the various ideas into one AIED system in order to optimize learning.
3. Few studies have focused on affect modulation strategies in AIED for learning.

Consequently, even if affective state detection is correctly implemented, the question of what an AIED system should do with this information remains unclear. In order to address this issue, this paper reviews the current literature of affect detection, of effects of affect on learning, and of affect modulation for learning optimization. We jointly discuss how this research can influence the design of AIED systems.

Several different affect frameworks have been used in previous studies [31, 38]. We chose to utilize Scherer’s affect framework [44] as a way to integrate the research findings pertaining to affect.

Scherer’s affect framework as well as the reasons we chose to use it are discussed in the first section. Section two presents the various approaches to automatic affect

1 In this paper, “affective state” is related to the specific and internal affective situation of an individual at a given moment, and can refer to several coexisting affective phenomena. Authors tend to agree that each affective phenomena is a process (rather than a state) that develops among multiple components [31, 44].
detection. Section three reviews the research on the effects of affect on learning and its related cognitive processes. Finally, section four discusses the existing affect-related strategies to ensure quality learning and places them in an AIED context.

1. Discussing Definitions of Affects: An Introduction to Scherer’s Framework

Affect-related words are difficult to define for a number of reasons. First, they are frequently used in daily interactions, “emotion”, “feeling”, “mood” and “attitude” being used interchangeably [33, 41, 44]. Natural language usage of affect-laden words does not always convey the full spectrum of an affective experience and may lead to overgeneralization [47]. Given this ambiguity, the scientific analysis of affect is difficult.

Furthermore, the scientific study of affect has revealed both cultural and contextual differences in affect. Affective antecedents (stimulus elements), subjective experience, appraisals, behavioural responses, and even physiological changes related to an emotional experience are reported to differ across cultures [36]. Indeed, the contextual nature of affective phenomena has been given little attention [36, 41].

There are two competing approaches to the study of affect. The first one is based on a categorical representation of discrete states [15, 38] and assumes that the words we use to describe a wide range of affective experiences are used consistently between and within individuals. The other approach relies on a dimensional representation of affective experiences [17, 42]. It assumes that affect can be broken down into a number of dimensions, valence and arousal being of primary focus [42]. Interoperability of these approaches is sometimes hard to achieve.

Scherer’s meta-analysis of affect provides a more coherent framework for comparing research on affect. In this framework, affective phenomena are described as component processes [44] that are interdependent and dynamic in that they can evolve over time. Scherer identified five affective features and linked them to a set of organismic subsystems arising from four different body systems that exist in every human being. Each of the affective features is responsible for fulfilling a specific affective function as shown in Table 1.

Table 1. Scherer’s list of organismic subsystems (adapted from [44])

<table>
<thead>
<tr>
<th>Organismic Subsystems</th>
<th>Associated Body System(s)</th>
<th>Affective Function</th>
<th>Related Affective Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information processing</td>
<td>CNS</td>
<td>Evaluation of objects and events</td>
<td>Appraisal of events (Cognitive arousal)</td>
</tr>
<tr>
<td>Support</td>
<td>CNS, NES, ANS</td>
<td>System regulation</td>
<td>Psychophysiological changes (neurophysiological component)</td>
</tr>
<tr>
<td>Executive</td>
<td>CNS</td>
<td>Preparation and direction of actions</td>
<td>Action tendencies (motivational component)</td>
</tr>
<tr>
<td>Action</td>
<td>SNS</td>
<td>Communication of reaction and behavioral intention</td>
<td>Motor expressions (face, voice, gesture)</td>
</tr>
<tr>
<td>Monitor</td>
<td>CNS</td>
<td>Monitoring of internal state and organism-environment interaction</td>
<td>Subjective experiences (feelings)</td>
</tr>
</tbody>
</table>


Scherer’s framework also clearly disambiguates six kinds of affective phenomena that are frequently confused in the literature (Table 2), using seven additional design features for such purpose (i.e., event focus, appraisal driven, response synchronization, rapidity of change, behavioural impact, intensity, duration).
Table 2. Scherer’s list of identified affective phenomena and their respective definitions (deduced from [44])

<table>
<thead>
<tr>
<th>Affective Phenomena</th>
<th>Definitions</th>
</tr>
</thead>
</table>
| **Emotions** | - “An episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism”  
- Scherer emphasizes the distinction between emotions i.e. “the total multimodal component process” and feelings i.e. “a single component [of any affective phenomena] denoting the subjective experience process” |
| **Moods** | - “Diffuse affect states, characterized by a relative enduring predominance of certain types of subjective feelings that affect the experience and behavior of a person”  
- “Often emerge without apparent causes”  
- “Generally of low intensity” |
| **Preferences** | - “Relatively stable evaluative judgments in the sense of liking or disliking a stimulus, or preferring it or not over other objects or stimuli” |
| **Attitudes** | - “Relatively enduring beliefs and predispositions towards specific objects”  
- “Can be labeled with terms such as hating, valuing or desiring” |
| **Affect Dispositions** | - “Tendency of a person to experience certain moods more frequently or to be prone to react with certain types of emotions” |
| **Interpersonal Stance** | - “Affective style that spontaneously develops, or is strategically employed in the interaction with the person or a group of persons”  
- “Examples: being polite, distant, cold, warm, supportive, contemptuous”  
- “Often triggered by events (encounters of a person), but less shaped by spontaneous appraisal than by affect dispositions, interpersonal attitudes, and most importantly strategic intentions” |

Scherer’s work is particularly valuable for formalizing affect-related research because, first, it is “genuinely interdisciplinary” [44] in that it includes major ideas, discoveries and findings concerning the nature of affect from many disciplines in the humanities and social sciences. Second, it promotes the objective study of the complex nature of affect. It resembles formal ontologies research in that the identity of a concept refers more to its essential and objective structure than to labels or symbol-based representations. Third, it distances itself from folk definitions by providing unambiguous and concise definitions that distinguish affective phenomena, and connect them to biological body systems of individuals. Fourth, it allows researchers to discuss dimension-based as well as category-based results at a meta-level. Fifth, it helps researchers approach intercultural differences in the affective domain [4, 36]. Lastly, it considers both trait-like affective phenomena, referring to “a general way of responding to the world, which varies by person, but is relatively stable” [33] (preferences, attitudes, affect dispositions, interpersonal stances) and state-like affects, reflecting “a response to the changing environment that is based on the situation and is less stable over time” [33] (emotion, mood).

2. Overview of Research on Automatic Affect Detection

Qualitative methods are frequently used to detect affect, such as Emote Aloud [9], Think Aloud, and interviews [33]. However, they require human intervention and, thus, are beyond the scope of this paper since our review is of automatic affect detection methods which address several of the affective features introduce in part 1.

Electroencephalography (EEG) [3, 28] refers to detecting neural electrical activity, using sensors strategically located on the head. Electromyography deals with the electrical activity of muscles. Unconscious facial muscles activation is notably used to detect cues of basic emotion occurrence [28, 29, 30, 34]. Galvanic Skin Response (GSR) measures the electrical resistance of the skin, which may evolve according to stress or arousal, for instance [3, 29, 34]. Classic GSR setups generally use two sensors located on two separate parts of the body (like two fingers that are far apart on the same
Cardiovascular Activity refers to several signals, such as heart rate or blood pressure that reflect heart activity, which may be highly affect-sensitive [29, 34]. Gradual fluctuations of the temperature of an individual can also be cues of mid-term affective phenomena [29]. Gesture Analysis refers to the analysis of body movements (arms, legs), most of the time by analyzing data obtained from a digital camera [48]. Posture analysis refers to the analysis of the evolution of the posture of an individual over time, frequently done using pressure sensors [13]. Facial recognition refers to the analysis of facial cues, mostly by referring to the Facial Action Coding System [15]. Analysis of various features of vocal interactions can also lead to affective diagnosis [48, 49]. Affective phenomena can also be addressed through analysis of several features of classic human-computer interactions [29], such as dialogue with a pedagogical agent [10]. Finally, because it is easy to set up, the use of questionnaires remains the most widespread affect detection method [7, 45]. By nature, questionnaires mostly address the subjective aspect (feeling) of affective phenomena [44].

Furthermore, when particularly considering specificities of those technical methods, it appears that features such as time latency, reliability and accuracy tend to disqualify the use of some of these methods for the assessment of specific affective phenomena. For instance, EEG analyzes patterns are very brief, and at first sight are not well indicated for assessing medium-to-long term phenomena such as mood. Scholars wishing to work on affective detection have to carefully consider the affective phenomena they want to address and the available data channels in order to select an appropriate method. However, sometimes, data processing techniques and algorithms may also widen the interest of a particular detection method. Hence, establishing the frequency of occurrence of positive or negative EEG patterns over a certain period of time could lead to an indirect measure of mood.

3. Affects in Cognition and Learning

This section reviews the interplay of affect and learning-related cognitive processes.

**Judgement.** Pre-existing moods influence how we judge current events [47]. Positive moods lead to higher ratings of tasks [16], social relations [18] and self-performance. Negative moods lead to negative judgements. For instance, angry people are more likely to blame human agents [26]. It would be interesting to investigate whether, according to the persona effect [32], there could be similar tendencies of angry learners towards pedagogical agents.

**Interest, attention, and learning.** Following Scherer’s definitions, interest could refer either to an emotion that is “activated” by an occurring activity or to an attitude toward an activity. Hirt et al. [23] suggest that participants who are in a positive mood before working on a task view the task as more interesting than individuals in a negative or neutral mood. This should lead to better performance and deeper satisfaction along with greater effort expenditure. Interest may lead to sustained attention or flow within an activity, and is negatively correlated to task-irrelevant thinking [39]. Conversely, negative affects, such as boredom, that stem from negative achievement, anxiety, shame, and hopelessness, negatively relate to flow experience and positively to task-irrelevant thinking [39].

**Motivation.** Individuals in a positive mood may be more apt to initiate problem-solving strategies. They have greater optimism regarding their resources and are less likely to anticipate significant barriers than those with negative moods who set higher
goals and are less optimistic about attaining them, resulting in decreased actions [46]. Positive attitudes towards a learning task as well as positive emotions while performing it, i.e. enjoyment of learning, are known to strengthen both intrinsic and extrinsic motivation, whereas negative attitudes and deactivating emotions, such as boredom and hopelessness, hinder motivation [24].

Memorization. Some researchers suggest a connection between mood at learning and at retrieval [16], but Kenealy [27] argues that this effect occurs only when there are no other recall cues. Negative affect narrows one’s thinking about details and analytic processing, which could favour tasks requiring rigid thought like rote memorization [1]. Goleman [20], however, suggests that anxious or depressed learners are unable to adequately assimilate information, due to the negative impact on the working memory.

Problem solving and creativity. A positive mood leads to more inclusive forms of thinking [1]. It has a direct influence on the quality of responses one gives [14] and an indirect impact on their quantity. Being in a sad mood may result in creative problem solving [26, 46]. Overall, moods do affect the decisions that individuals make throughout the problem solving process, all the way from the (usually neglected) problem detection stage to the acceptance of the solution [47].

Information processing. Individuals who are happy are more apt to employ a heuristic processing strategy that is marked by top-down processing; they count heavily on pre-existing knowledge structures and focus minimally on the details of the situation. In contrast, those who are in a sad mood are more likely to utilize a systematic processing strategy that is marked by bottom-up processing; they depend little on pre-existing knowledge structures, instead paying close attention to the details at hand [47]. Indeed, being in a negative affective state is linked to a more narrow focus of attention [6] and an increased level of spontaneous causal reasoning.

Decision making. Individuals experiencing a sad mood are profoundly influenced by strong arguments and not moved by weak ones, while happy individuals are moderately, but equally influenced by both types of arguments [47]. Individuals in a sad mood report fewer inconsistencies in multi-attribute decision tasks than do their happy counterparts [47], and while engaging in a chicken game, they are likely to base their decisions on a rational analysis of the game’s structure [22]. When engaged in a similar game, individuals in a happy mood tend to heuristically imitate the decisions of other players.

4. Affect Management for Ensuring Quality Learning in AIED

Finally, this paper explores strategies that may be used to promote a computer-supported learning task. They are discussed according to the affective phenomena they address. Preferences and affect dispositions are ignored, as they are rigid phenomena.

Strategies related to emotions.

Emotional Induction. Promoting positive emotions while engaged in a learning task may lead the learner to develop a positive attitude towards this task. This in turn results in an increased willingness to get involved in similar activities in the future. Providing learners with affective antecedents (events or objects with an affective charge), such as movies or pictures [29], is a common method of emotional induction. However, research has shown that emotional induction is completely successful only when participants are not aware that their emotions are being manipulated and hence do not explicitly focus on them [5].
Emotional Suppression. Emotions are intense affective phenomena [44] that involve almost all body systems, which can lead to focusing on the affective antecedent (whether positive or negative), thus disrupting the learning process [44]. Potentially, learning could be enhanced by reducing or suppressing emotions. Expressive suppression is a cognitive method dealing with this, in which outward expressions of emotions are suppressed during social interaction [21, 50]. Cognitive reappraisal is another existing method, in which one is trained to monitor his/her thoughts, by finding negative ones and replacing them with more positive ones [31].

Preferences and attitudes towards the learning content should also be monitored to determine when suppressive or inductive emotional strategies should be employed.

Strategies related to mood.

Mood Induction. The recall of a happy or a sad event, followed by a period of writing related to the event, has been shown to be an efficient method for inducing mood [5], and produces better results when using sad events. Methods to induce moods should be as transparent as possible in order to optimize the learning benefits. Alternative methods, such as biofeedback method, which teaches learners to manage their brain activity by showing them real-time representation of it [12], could also be explored. Similarly, music is also a well-known mood inducer, but learners’ musical tastes must be taken into account.

Strategies related to attitudes.

Promoting positive attitude towards the activity. Negative attitudes such as computer anxiety may restrain the willingness of a student to get involved in a computer-supported learning activity. This could be addressed by considering learners’ self-efficacy beliefs [2]. As seen earlier, emotional induction methods could also lead one to associate the learning activity with self-satisfaction emotions. Another strategy would be to develop learners’ usefulness appreciation of the activity by showing them how this helps them achieve their personal goals [35]. The approach used to present learning content must also be taken into account given that several studies stress the affective interest of good game design [19, 43].

Promoting ethos of the system. Ethos (credibility) is a core element in the rating of information systems and technology-mediated practices, and varies according to the cultural and personal values of users (or learners). Such elements should be considered when designing the interface of an AIED system [37], given that bad ethos may cause a lack of trust toward the system.

Strategies related to interpersonal stances.

Promoting positive interpersonal stances within AIED systems has not been considered in great detail, even if, according to the persona effect [32], similar positive affect could be obtained while interacting with a computer and with a real human. Johnson et al. [25] have shown that endorsing a polite attitude could help a pedagogical agent to better communicate with learners. Positive interpersonal stances could also be obtained by interacting with learning companions or other individuals. Strategies to manage interpersonal stances during collaborative activities should also be considered.

5. Conclusion

In this paper we reviewed the diverse research on affect as a first step prior to developing an AIED system that could detect and then intelligently manage the affective dimension of the learner. We introduced Scherer’s framework as an extremely
useful tool to disambiguate the definitions of various affective phenomena as well as their interrelations with each other. During our review of automatic affect detection methods, we realized that some methods were more significant for the detection of some affective phenomena than for others. However, as mentioned in [41], the ability to detect affect by a tutor has very limited value without giving appropriate feedback. Thus, in section three we reviewed the interplay which exists between the cognitive aspects of learning and affect. We found that mood has the strongest link to learning, while other affective phenomena have a smaller but still important relationship with learning. Following this, we reviewed the major affect-related strategies which might be useful in promoting quality learning.

In conclusion, it is clear that emotions and moods should be considered carefully if one wishes to develop a truly intelligent and adaptive tutor. Moreover, other affective phenomena could create interesting opportunities for affective support that may have been somewhat neglected by the AIED community. An example of this would be the investigation of interpersonal stances and related sociocultural influences between all actors of social and participative learning software [4, 25].

The current state-of-the-art technology brings us marvelously close to being able to automatically detect the affective states of a student. Likewise, we see strong evidence linking affect to cognitive processing. Additionally, we possess a large set of strategies which can be used to modulate affect. Only through integrating these research streams can we hope to develop an AIED system that can be said to be affect intelligent.

References