Predicting Facial Indicators of Confusion with Hidden Markov Models

Joseph F. Grafsgaard, Kristy Elizabeth Boyer, and James C. Lester

Department of Computer Science, North Carolina State University
Raleigh, North Carolina, USA
{jfgrafsg, keboyer, lester}@ncsu.edu

Abstract. Affect plays a vital role in learning. During tutoring, particular affective states may benefit or detract from student learning. A key cognitive-affective state is confusion, which has been positively associated with effective learning. Although identifying episodes of confusion presents significant challenges, recent investigations have identified correlations between confusion and specific facial movements. This paper builds on those findings to create a predictive model of learner confusion during task-oriented human-human tutorial dialogue. The model leverages textual dialogue, task, and facial expression history to predict upcoming confusion within a hidden Markov modeling framework. Analysis of the model structure also reveals meaningful modes of interaction within the tutoring sessions. The results demonstrate that because of its predictive power and rich qualitative representation, the model holds promise for informing the design of affective-sensitive tutoring systems.

Keywords: Affect prediction, hidden Markov models, intelligent tutoring systems, tutorial dialogue.

1 Introduction

One-on-one human tutoring is highly effective for student learning [1]. Intelligent tutoring systems (ITSs) hold great promise for achieving this level of effectiveness, with many such systems producing significant learning gains [2,3]. Recent advances in ITS research such as modeling the strategies of expert human tutors [4] and examining learner emotions during tutoring [5] continue to enhance the effectiveness of ITSs. To date, a number of ITSs and tutorial dialogue systems have begun to address learner affect [6-8]. The emerging results highlight the importance of incorporating models of learner emotions into ITSs to provide students with more effective learning experiences.

Predicting student affect is an essential step toward identifying optimally effective behavior within affectively aware intelligent systems. A promising means for predicting student affect is through analyzing learners' facial expressions. Recent investigations have identified facial expression configurations that relate to learner emotions [5,9,10]. These results build on advances in human emotion modeling that are tied to particular cross-cultural facial expressions [11]. Compared to other modalities of affect detection, facial expressions may be a richer, more informative channel during learning [2,6]. Yet, the field is far from assembling a comprehensive

catalogue of learner emotions and the facial expressions with which they correlate [5,12].

Although no comprehensive catalogue of learner emotions currently exists, there is widespread agreement that occurrences of *confusion* are highly relevant during learning [5,9,13]. While identifying episodes of confusion presents significant challenges, a promising approach leverages student facial expression. Numerous studies of the correlations between facial expressions and learner emotions have identified a link between confusion and the facial *action unit 4*, which is the "Brow Lowerer" movement, referred to as *AU4* [5,9,10,14]. The present work utilizes these findings to predict student confusion as evidenced by student AU4 during computer-mediated human-human tutoring.

The modeling approach presented in this paper represents the tutoring session (consisting of dialogue, task actions, and facial expressions) as a sequential set of observations. The goal is to predict unseen observations based on the prior observations. While a number of sequential modeling techniques may be appropriate for this task, a particularly promising framework is the hidden Markov model (HMM), which has been successfully used to model tutorial strategies within a tutorial dialogue corpus [15,16]. The present findings demonstrate that HMMs can learn a predictive model of *confusion*, as indicated by student AU4, from a corpus of human tutoring.

2 Related Work

Recent tutoring research has identified a set of cognitive-affective states that are particularly relevant to learning: *anxiety*, *boredom*, *confusion*, *curiosity*, *delight*, *eureka*, *flow*, and *frustration* [9,17]. These learner emotions appear to have different effects on learning and motivation [5,18]. For example, frustration seems to be a problematic cognitive-affective state, which promotes a negative "state of stuck" [7]. While frustration may have negative consequences, an even more detrimental state is boredom, which severely inhibits student learning [18]. In both cases, a paramount concern is the persistence of negative affective states, which can lead to a "vicious cycle" [5,8].

While the cognitive-affective state of confusion may at first glance appear to be negative, it has increasingly been shown to coincide with moments of learning [5,8]. The positive impact of confusion may be due to the associated concept of cognitive disequilibrium, which involves a moment of uncertain knowledge that is (ideally) subsequently revised to reach correct understanding [13]. Thus, confusion may function as an essential intermediary state on the path of deep learning [2,13]. This notion has been supported in studies across multiple learning environments [2,8,18].

Effectively incorporating affect in intelligent systems requires the capability to diagnose instances of learner emotion [19]. This diagnosis involves both detecting and understanding emotions, and recent years have seen increased research into both problems [2,5,20]. Facial expressions are a particularly meaningful channel for both learner affect detection and understanding [5,6,12]. With respect to affect detection, manual and automated recognition of facial expressions have been shown to improve predictive models [6,21,22]. With respect to affect understanding, specific facial configurations are known to be associated with learner affect. Confusion has been

correlated with facial action unit 4 (AU4, "Brow Lowerer") in multiple studies, based on self, peer, and FACS-certified expert judgments of affective events [5,9,10].

A common thread in work on affect detection and understanding is the importance of identifying how learner affect follows from context. For this reason, a predictive model of affect holds great promise, not only for influencing the behavior of affectively aware ITSs, but also for informing a fundamental understanding of emotions during learning. This paper presents a predictive model of student confusion created using hidden Markov models (HMMs). The findings indicate that by leveraging dialogue, task, and facial expression history, HMMs can predict the presence of student AU4. These results have implications both for fundamental investigations of learner emotions, and for predicting affect during tutoring.

3 Corpus and Manual Annotations

A corpus of human-human tutorial dialogue was collected during a tutorial dialogue study [15]. Students solved an introductory computer programming problem and engaged in computer-mediated textual dialogue with a human tutor. The corpus consists of 48 dialogues annotated with dialogue acts (Table 1). Annotations also include information about student progress on the programming task [16].

Table 1. Dialogue act tags and frequency in corpus (*S* = student, *T* = tutor)

Act	Description	*S*	*T*
ASSESSING QUESTION	Task-specific query or feedback request	44	83
EXTRA DOMAIN	Unrelated to task	37	42
GROUNDING	Acknowledgement, thanks, greetings, etc.	57	38
LUKEWARM CONTENT FDBK	Partly positive/negative elaborated feedback	2	23
LUKEWARM FEEDBACK	Partly positive/negative task feedback	3	21
NEGATIVE CONTENT FDBK	Negative elaborated feedback	5	77
NEGATIVE FEEDBACK	Negative task feedback	10	10
POSITIVE CONTENT FDBK	Positive elaborated feedback	10	21
POSITIVE FEEDBACK	Positive task feedback	23	119
QUESTION	Conceptual or other query	31	24
STATEMENT	Declaration of factual information	55	320

Student facial video was collected during the tutoring sessions, but the videos were not shown to tutors. Fourteen of these videos were annotated with student displays of AU4 (Figure 1) using the Facial Action Coding System (FACS) [1] [14,23]. One certified FACS coder annotated all fourteen videos from start to finish, pausing at all observed instances of AU4. A second certified FACS coder annotated a subset of six videos. The continuous intervals of time were discretized into one-second intervals, on which intercoder agreement was κ=0.86 (Cohen's kappa). Annotated excerpts of the corpus are shown in Figure 2, which also displays the best-fit sequence of hidden states as identified by the HMM (Section 5.1).

[1] Manual annotation of facial action units is very labor intensive. Comprehensive FACS coding typically requires at least sixty hours per hour of video. Annotating a subset of AUs is faster, requiring approximately ten hours per hour of video.

Figure 1. Student displays of AU4

Excerpt 1			
13:16:03	Tutor:	no, it's easier than that, you just have to make the middle if into an "else if" [NEGATIVE CONTENT FEEDBACK]	STATE 10
	Student:	CORRECT TASK ACTION AU4	STATE 6
13:16:31	Tutor:	does that make sense? [ASSESSING QUESTION AU4]	STATE 10
13:16:41	Tutor:	that way it only checks the 2nd conditional if the first one failed [STATEMENT]	STATE 8
13:17:20	Student:	it makes sense now that you explained it […] [POSITIVE CONTENT FEEDBACK]	STATE 4
Excerpt 2			
14:52:18	Tutor:	no, before we start sorting [NEGATIVE CONTENT FEEDBACK AU4]	STATE 10
	Student:	CORRECT TASK ACTION AU4	STATE 6
14:52:27	Tutor:	so, before the first loop you can use i for this loop counter if you want to [STATEMENT AU4]	STATE 10
	Student:	MIXED PROGRESS TASK ACTION AU4	STATE 6
14:53:52	Student:	i try to keep them different so i don't confuse myself [STATEMENT]	STATE 7

Figure 2. Excerpts from annotated tutoring session corpus, with most probable sequences of HMM hidden states (Section 5)

Figure 3 shows the frequencies of AU4 corresponding to tutor and student dialogue acts. For tutor dialogue acts, an instance of AU4 is considered "corresponding" if it occurs within ten seconds after the tutor move; for student acts, ten seconds before the student move. These durations were empirically determined to account for student preparation of an utterance and reception of a tutor utterance. Of student utterances, LUKEWARM CONTENT FEEDBACK corresponds to the highest probability of student AU4. In this dialogue move students articulate partially correct knowledge. Of tutor utterances, the most likely to correspond to student AU4 is NEGATIVE FEEDBACK, in which the tutor states that the student has made a mistake but does not provide an explanation. Another dialogue move that has a relatively high probability of AU4 is student ASSESSING QUESTION, which constitutes a direct request for task-based feedback. Students generally make these requests when their confidence in a recent task action is low.

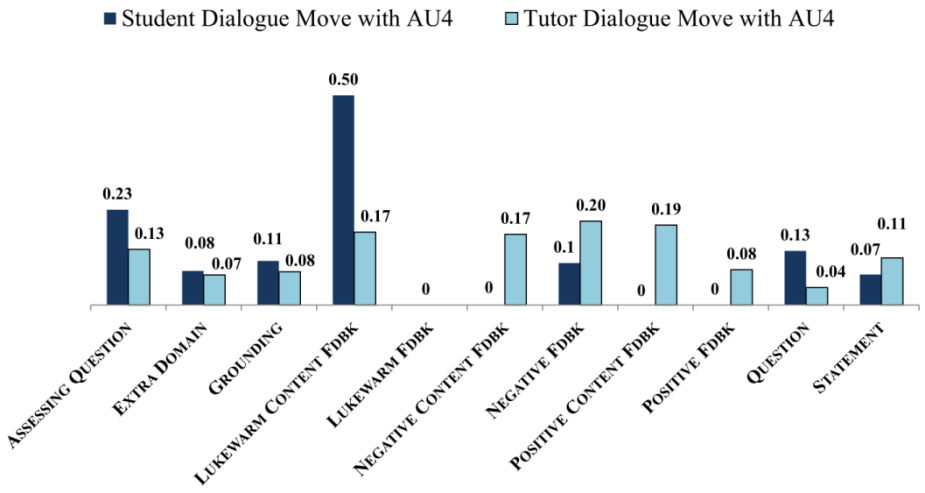

Figure 3. Relative frequency of student and tutor dialogue moves with AU4

Task actions were labeled based on progress toward a correct solution to the programming problem at hand, at a between-dialogue-moves granularity. Each task action cluster was characterized as CORRECT, INCOMPLETE, INCORRECT, or MIXED PROGRESS (a mixture of correct, incomplete, and/or incorrect task actions). As shown in Figure 4, students most frequently displayed AU4 during episodes of MIXED PROGRESS (37% of the time). Students were less likely (24%) to display AU4 during episodes of INCORRECT task action. As novices, these students were likely unaware of their mistakes when undertaking a completely incorrect task action. On the other hand, partially correct and partially incorrect task episodes indicate the student had sufficient knowledge to recognize that errors were present, and may have been experiencing constructive confusion toward reaching increased understanding.

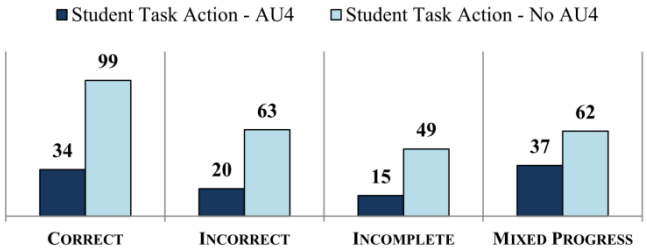

Figure 4. Frequency of student task actions with AU4 present or absent

4 HMM Learning and Prediction of Student AU4

A hidden Markov model (HMM) is defined by an *initial probability distribution* across hidden states, *transition probabilities* among hidden states, and *emission probabilities* for each hidden state and observation symbol pair [24]. The hidden states represent the underlying probabilistic system that generates a given sequence of observed events. The initial state probability gives the possibility of beginning in any

hidden state. Transition probabilities encode the likelihood of entering one hidden state from another. Emission probabilities encode the likelihood of producing a given observation from a particular hidden state. HMMs learn statistical dependencies between hidden states and the corresponding observations. The hidden state structure can then be analyzed to identify underlying trends. Using HMMs, it is possible to uncover a rich interplay between learner affect, tutorial dialogue and task context.

4.1 Model Learning

The observation sequences consist of annotated observations from the corpus, including dialogue moves by tutor or student, or student task action segments. Each of these observations also includes a tag for whether student AU4 was associated with that event. For example, the observation symbol sequence that corresponds to Excerpt 1 in Figure 2 is, [STUDENT NEGATIVE CONTENT FEEDBACK NOAU4, CORRECT TASK ACTION AU4, TUTOR ASSESSING QUESTION AU4, TUTOR STATEMENT NOAU4, STUDENT POSITIVE CONTENT FEEDBACK NOAU4].

The HMMs were learned within a leave-one-out framework. Within each fold, five random restarts of model parameters were performed to reduce the potential of model convergence at a local optimum. An additional outer training loop, ranging from two to twenty, was performed to identify the optimal number of hidden states. The best-fit model has eighteen hidden states, and its structure is discussed in detail in Section 5.

4.2 Prediction of AU4

The leave-one-out design resulted in fourteen training/testing folds, one for each tutoring session. Four of these sessions contained an observation symbol (combination of dialogue move and AU4 presence/absence) that occurred nowhere else in the data, so the learned model was not used to predict on these sessions. Predictive findings from the remaining ten test sessions are presented here, though an online predictive model used during tutoring could address this by learning across all possible symbols in the state space, regardless of absence in a particular session. The predictive accuracies of the HMMs are compared against a majority class baseline as well as a first-order observed Markov model (OMM) (Table 2).

Table 2. Comparison of predictive accuracy of classifiers; accuracies that are statistically significantly better than baseline are in bold (paired *t*-test, $p < 0.005$)

Classifier	Accuracy (across sessions)	Std. Dev. of Accuracy
HMM Train	**0.868**	0.021
HMM Test	**0.907**	0.059
OMM Train	0.186	0.013
OMM Test	0.557	0.284
Baseline	0.845	-

OMMs do not include hidden states, and thus condition the present state purely on the transition probability distribution from the previous state. The training set predictive accuracies indicate that HMMs fit the training data better than the other models. The predictive accuracy of the learned HMMs on the test set was higher on average than

predictions on the training set, but not surprisingly, the standard deviation was also greater. Both the training and test predictions greatly outperformed the predictive accuracy of the OMMs, which were below baseline. This below-baseline performance of OMMs indicates that the presence or absence of AU4 at time *t* is not highly predictive of the presence or absence of AU4 at time *t+1*, which is an interesting discovery in this corpus. However, the additional stochastic structure provided by the HMM is able to predict AU4 significantly above baseline.

5 Discussion

The predictive accuracies of the HMMs suggest that these models hold great promise for learner affect prediction. On unseen test data, the HMMs predicted significantly better than an OMM and a (very high) majority class baseline. To gain more insight into how the HMM structure facilitates prediction of student AU4, this section examines and interprets the structure of the learned (best-fit) HMM.

5.1 Hidden State Structure

HMMs' predictive power is gained in part by the way these models can learn higher-order structure (in the form of hidden states) based on observation sequences. The model structure shown in Figure 5 illustrates this, with emission probability distributions displayed as bar graphs and transitions as edges. To facilitate discussion, the states were named after model learning through qualitative analysis. STATE 6, *Student Work with Confusion,* is dominated by student task actions with AU4 present. STATE 10, *Tutor Help*, emits a combination of tutor dialogue moves with AU4 present, and tutor feedback with no AU4. STATE 4, *Overcoming Confusion*, is dominated by tutor statements with AU4 present and student positive feedback. This state corresponds to tutor statements that are not consistent with students' prior knowledge. Interestingly, the state also generates student positive feedback, which may indicate that students moved past cognitive disequilibrium and into a state of understanding. STATE 16, *Conversational Grounding*, primarily encompasses non-task-oriented student and tutor dialogue moves, but also generates with small probability student negative feedback without AU4.

These emission probability distributions indicate ways in which the HMM abstracts from observation sequences to meaningful higher-order structure. Of equal importance are the transition probabilities between the hidden states. STATE 6 and STATE 10 are more likely to transition to each other than any other hidden states. This transition is illustrated in one sequence of events (Figure 2 Excerpt 1) in which the tutor provides negative content feedback followed by a student correct task action with confusion present (as evidenced by AU4). The tutor then asks an assessing question to gauge the student's understanding, which also coincides with a moment of confusion. The tutor further explains the computer programming concept with an instructional statement clarifying prior feedback. The student then takes a moment to reflect on the material and informs the tutor that the explanation was helpful. This example demonstrates the strong connection between *Student Work with Confusion*, STATE 6, and *Tutor Help*, STATE 10. Meaningful tutor feedback and instruction that induce confusion are produced in STATE 10, while STATE 6 corresponds with student tasks actions accompanied by confusion. Both states are highly relevant to learning.

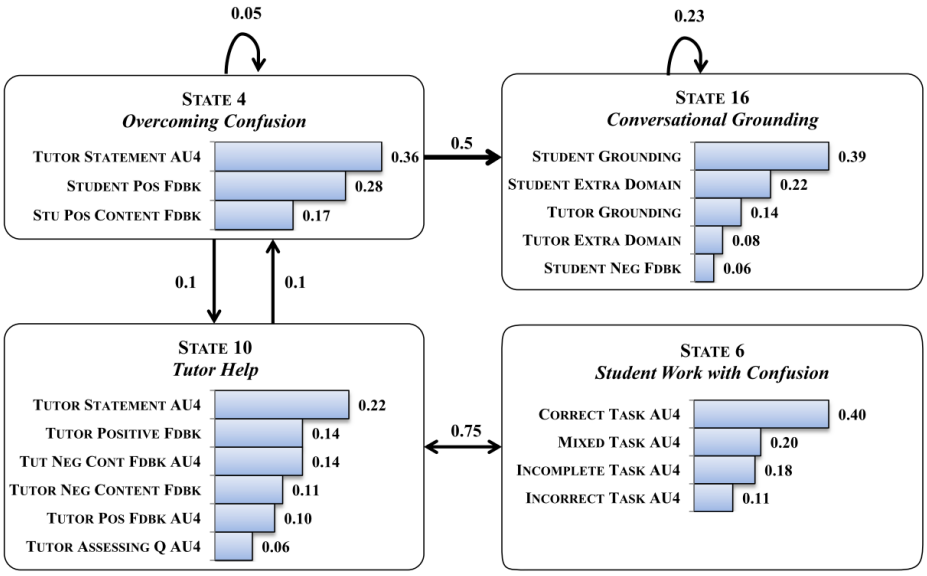

Figure 5. Learned HMM structure: a subset of four hidden states is shown, with transition (arrow) and emission (bar chart) probabilities ≥ 0.05 indicated

The second example (Figure 2 Excerpt 2) further characterizes the interplay of STATE 6 and STATE 10, with the student progressing on the programming task (with AU4 present) while receiving tutor feedback and instruction. The excerpt begins with tutor negative content feedback on the student's current task progress, with student confusion indicated by AU4. The student then immediately completes the subtask, still showing AU4. The tutor continues instructing the student with a comment on a relevant programming concept (AU4 still present). The student then continues programming, with mixed progress (AU4 continues). After approximately a minute of working on the task, the student responds to the tutor statement with an explanation of the work performed. Thus, the student displays confusion until after the tutor completes instruction. Both excerpts seem to show effortful learning, with a combination of instruction during *Tutor Help* and task progress in *Student Work with Confusion*. Therefore, HMMs represent a promising approach to automatically learn semantically meaningful affect-rich models of tutorial interaction.

5.2 Limitations

There are two primary limitations that should be noted. First, manual annotation of facial expressions is very labor intensive, which constrained the number of sessions analyzed for learner affect. Automated techniques for FACS coding [14] are actively being investigated, although they are not currently as accurate as manual annotation [21,25]. Second, a learned HMM may potentially require large amounts of data to produce a predictive model generalizable enough to deploy within a larger population, as evidenced by the observation symbol sparsity that was encountered in this analysis. Further studies are necessary to evaluate the generalizability of predictive HMM models of learner affect and their use in online prediction during tutoring.

6 Conclusion

Learner affect plays a vital role in the success or failure of a tutorial interaction. In particular, the cognitive-affective state of confusion is highly relevant on the path to acquiring knowledge since confusion accompanies learning impasses during which students must resolve misconceptions that challenge their conceptual understanding. Predicting confusion is an important step toward understanding the effects of various ITS interventions and toward designing more effective strategies.

This paper has presented a novel predictive model of learner confusion that incorporates dialogue moves, task performance, and facial expression using hidden Markov models (HMMs). Such models may play an important role in the diagnosis of learner confusion for future systems. Additionally, analysis of the model structure identified meaningful transitions between affect-enriched states of tutor and student dialogue moves and student task progress. In future work, affect-predictive HMMs need to be further developed by incorporating data regarding a wider set of learner affective states. These future predictive models may be instrumental in diagnosing learner affect during interactions with intelligent tutoring systems.

Acknowledgements

This work is supported in part by the NC State University Department of Computer Science along with the National Science Foundation through Grants IIS-0812291, DRL-1007962 and the STARS Alliance Grant CNS-0739216. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

References

1. Bloom, B.S. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. Educational Researcher. 13, pp. 4-16 (1984).
2. D'Mello, S.K., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., Perkins, L., Graesser, A.C. A Time For Emoting: When Affect-Sensitivity Is and Isn't Effective at Promoting Deep Learning. Proceedings of 10th International Conference on Intelligent Tutoring Systems. p. 245–254 (2010).
3. Koedinger, K.R., Anderson, J.R., Hadley, W.H., Mark, M.A. Intelligent Tutoring Goes To School in the Big City. Intl. Jl. of Artificial Intelligence in Education. 8, 30-43 (1997).
4. D'Mello, S., Olney, A., Person, N. Mining Collaborative Patterns in Tutorial Dialogues. Jl. of Educational Data Mining. 2, 1-37 (2010).
5. D'Mello, S.K., Lehman, B., Person, N. Monitoring Affect States During Effortful Problem Solving Activities. International Jl. of Artificial Intelligence in Education. 20, (2010).
6. Arroyo, I., Cooper, D.G., Burleson, W., Woolf, B.P., Muldner, K., Christopherson, R.M. Emotion Sensors Go To School. 14th International Conference on Artificial Intelligence in Education (2009).
7. Burleson, W. Affective Learning Companions: Strategies for Empathetic Agents with Real-Time Multimodal Affective Sensing to Foster Meta-Cognitive and Meta-Affective Approaches to Learning, Motivation, and Perseverance. MIT Ph.D. thesis, (2006).
8. McQuiggan, S.W., Robison, J.L., Lester, J.C. Affective Transitions in Narrative-Centered Learning Environments. Educational Technology & Society. 13, 40–53 (2010).

9. Craig, S.D., Graesser, A.C., Sullins, J., Gholson, B. Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. Jl. of Educational Media. 29, 241–250 (2004).
10. McDaniel, B.T., D'Mello, S.K., King, B.G., Chipman, P., Tapp, K., Graesser, A.C. Facial Features for Affective State Detection in Learning Environments. Proceedings of the 29th Annual Meeting of the Cognitive Science Society. p. 467–472 (2007).
11. Russell, J.A., Bachorowski, J., Fernandez-Dols, J. Facial and vocal expressions of emotion. Annual Review of Psychology. 54, 329–49 (2003).
12. Afzal, S., Robinson, P. Natural Affect Data - Collection & Annotation in a Learning Context. Proceedings of the International Conference on Affective Computing and Intelligent Interaction. pp. 1-7 (2009).
13. Graesser, A.C., Olde, B.A. How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. Jl. of Educational Psychology. 95, 524–536 (2003).
14. Ekman, P., Friesen, W.V., Hager, J.C. Facial Action Coding System. A Human Face, Salt Lake City, USA (2002).
15. Boyer, K.E., Ha, E.Y., Wallis, M., Phillips, R., Vouk, M., Lester, J. Discovering Tutorial Dialogue Strategies with Hidden Markov Models. Proceedings of the 14th International Conference on Artificial Intelligence in Education. p. 141–148 (2009).
16. Boyer, K.E., Phillips, R., Ingram, A., Ha, E.Y., Wallis, M.D., Vouk, M.A., Lester, J.C. Characterizing the Effectiveness of Tutorial Dialogue with Hidden Markov Models. Proceedings of the 10th International Conference on Intelligent Tutoring Systems. p. 55–64 (2010).
17. D'Mello, S.K., Craig, S.D., Graesser, A.C. Multi-Method Assessment of Affective Experience and Expression during Deep Learning. International Jl. of Learning Technology. 4, 165–187 (2009).
18. Baker, R.S.J. d, D'Mello, S.K., Rodrigo, M.M.T., Graesser, A.C. Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments. International Jl. of Human-Computer Studies. 68, 223–241 (2010).
19. Picard, R.W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D., Strohecker, C. Affective Learning — A Manifesto. BT Technology Jl.. 22, 253–269 (2004).
20. Woolf, B.P., Burleson, W., Arroyo, I., Dragon, T., Cooper, D.G., Picard, R.W. Affect-aware tutors: recognising and responding to student affect. International Jl. of Learning Technology. 4, 129–164 (2009).
21. Calvo, R.A., D'Mello, S.K. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. IEEE Transactions on Affective Computing. 1, 18–37 (2010).
22. D'Mello, S.K., Graesser, A. Multimodal Semi-Automated Affect Detection from Conversational Cues, Gross Body Language, and Facial Features. User Modeling and User-Adapted Interaction. 20, 147-187 (2010).
23. Grafsgaard, J.F., Boyer, K.E., Phillips, R., Lester, J.C. Modeling Confusion: Facial Expression, Task, and Discourse in Task-Oriented Tutorial Dialogue. Proceedings of the 15th International Conference on Artificial Intelligence in Education. p. 98-105 (2011).
24. Rabiner, L.R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE. 77, 257 -286 (1989).
25. Zeng, Z., Pantic, M., Roisman, G.I., & Huang, T.S. A Survey Of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 31, 39-58 (2009).