Toward a Machine Learning Framework for Understanding Affective Tutorial Interaction

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 and cognition intertwine throughout human experience. Research into this interplay during learning has identified relevant cognitive-affective states, but recognizing them poses significant challenges. Among multiple promising approaches for affect recognition, analyzing facial expression may be particularly informative. Descriptive computational models of facial expression and affect, such as those enabled by machine learning, aid our understanding of tutorial interactions. Hidden Markov modeling, in particular, is useful for encoding patterns in sequential data. This paper presents a descriptive hidden Markov model built upon facial expression data and tutorial dialogue within a task-oriented human-human tutoring corpus. The model reveals five frequently occurring patterns of affective tutorial interaction across text-based tutorial dialogue sessions. The results show that hidden Markov modeling holds potential for the semi-automated understanding of affective interaction, which may contribute to the development of affect-informed intelligent tutoring systems.

Keywords: Affect, hidden Markov models, tutorial dialogue.

1 Introduction

Research in recent years has highlighted the interplay of cognition and affect in tutorial interaction. This interplay has implications for the design of intelligent tutoring systems (ITSs) that seek to attain or exceed the effectiveness of expert human tutors. To meet this goal, recent results demonstrated that understanding both the cognitive and affective nature of tutorial interaction may be necessary [1]. Affective phenomena during interactions with ITSs have been examined through a wide array of modalities including self-reports, observation, system logs, dialogue, facial expression, posture, and physiological measures [1]. Prior investigations of facial expression in tutoring identified links between particular facial movements and cognitive-affective states relevant to learning [2].

This paper details the construction and analysis of a descriptive HMM built from task-oriented textual tutorial dialogue annotated with dialogue acts and facial expression annotated from video. Facial movement combinations were annotated in a novel, three-phase protocol to provide rich affective representation within tutorial

S.A. Cerri et al. (Eds.): ITS 2012, LNCS 7315, pp. 52–58, 2012.

dialogue. Analysis of the learned HMM structure revealed five prevalent and persistent patterns of affective tutorial interaction represented by recurring sequences of hidden states. These results show the potential of HMMs for semi-automated understanding of affective tutorial interaction, which may inform integration of affect into future ITSs.

2 Related Work

Few studies have utilized hidden Markov models (HMMs) to model affect within the context of learning. In a recent study based on interactions with AutoTutor [3], HMMs learned transitions primarily consistent with the theory of cognitive disequilibrium. In an earlier study with Wayang Outpost [4], a math ITS for standardized test preparation, an HMM that modeled motivation improved predictive accuracy in a dynamic mixture model for correctness of student responses. Both approaches added constraints on top of those inherent within HMM assumptions. A recent study of human-human tutoring that modeled student brow lowering (an indicator of confusion) using HMMs provided both a predictive model and an analysis of confusion within the tutorial interaction [5]. The work presented here builds on these prior findings by leveraging sixteen facial movements (including brow lowering) in a purely descriptive model built without additional constraints, resulting in a richer representation of affect.

3 Dialogue Corpus and Facial Expression Annotation

A corpus of human-human tutorial dialogue was collected during a tutorial dialogue study [6]. 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, shown in Table 1. Student facial video was collected for post-analysis. (Note that the videos were not shown to tutors.) Seven of the highest quality facial videos were selected for the extent to which the student's entire face was visible during the recording, and for near-even split across genders and tutors. These videos were annotated with facial expressions for the present analysis (selected examples are shown in Figure 1). Tutoring sessions ranged in duration from thirty minutes to over an hour.

The seven selected facial videos were manually annotated using the Facial Action Coding System (FACS), which enumerates the possible movements of the face through a set of facial action units (AUs) [7]. Two certified FACS coders viewed entire videos, encoding facial events of one or more AUs with a start and end frame. Some FACS AUs were excluded due to excessive burden in manual FACS coding (e.g., mouth opening, blinking) or anticipated rarity (e.g., lip pucker, lip funneler). Sixteen were selected for coding: AUs 1, 2, 4-7, 9, 10, 12, 14-17, 20, 23, 24, and 31.

In the first phase of the condensed FACS protocol, the two certified FACS coders independently annotated occurrences of AUs. The coders met in a second phase to produce a combined set of facial event instances without discussing specific AUs, during which event instances were merged or eliminated. By the end of the second phase, the coders agreed completely upon the start and end time of facial events (without discussing specific AUs). In the third phase, one of the coders reviewed where the facial events occurred and decided on precisely which AUs occurred. Finally, the second coder annotated 9.3% of the facial events independently, establishing an agreement average of Cohen's κ=0.67, comparable with similar studies [2].

Table 1. Dialogue act tags and frequency across the seven sessions (*S* = student, *T* = tutor)

Act	Description	*S*	*T*
Assessing Question	Task-specific query or feedback request	16	29
Extra Domain	Unrelated to task	20	26
Grounding	Acknowledgement, thanks, greetings, etc.	26	16
Lukewarm Feedback	Partly positive/negative task feedback	2	12
Lukewarm Content Fdbk	Partly positive/negative elaborated feedback	1	9
Negative Feedback	Negative task feedback	5	5
Negative Content Fdbk	Negative elaborated feedback	1	34
Positive Feedback	Positive task feedback	10	76
Positive Content Fdbk	Positive elaborated feedback	2	5
Question	Conceptual or other query	13	9
Statement	Declaration of factual information	18	143

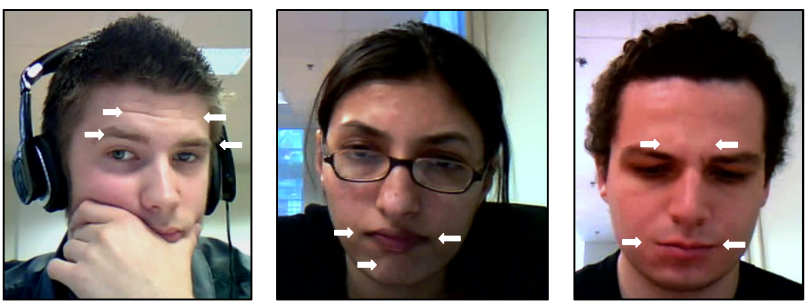


Fig. 1. Examples of facial action units: AUs 1+2 or "surprise" (left), 14+17 or "doubt" (center), and 4+12 or "confusion and frustration" (right). Arrows indicate facial movements.

This event-based annotation protocol incorporates AU combinations, which denote multiple facial movements occurring at the same time. While related research has indicated some facial expression and emotion correlations [2,7], affect-facial expression mapping is a difficult problem that requires considering the surrounding context. Affective interpretations discussed here are based on the simplified tutorial context offered by computer-mediated tutorial interaction.

4 Hidden Markov Modeling and Discussion

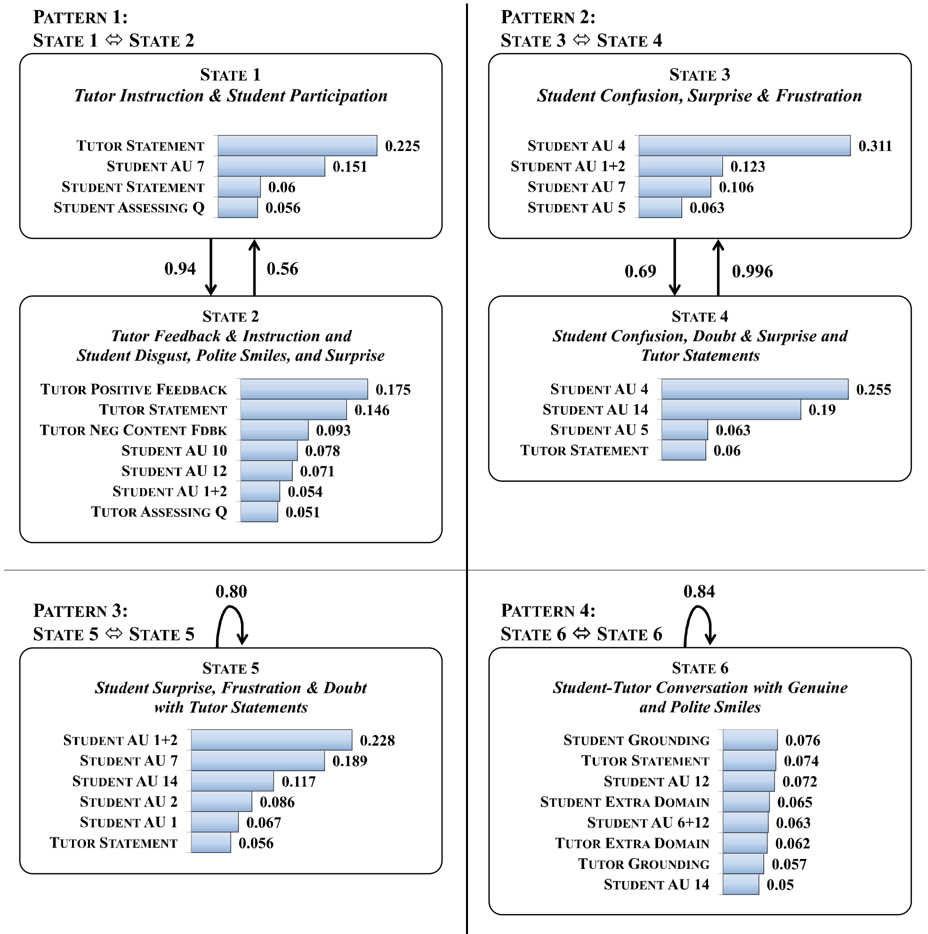
A hidden Markov model (HMM) is defined by an *initial probability distribution* across hidden states, *transition probabilities* between hidden states, and *emission probabilities* for each hidden state and observation symbol pair [8]. HMMs learn a probabilistic structure that preserves patterns within the modeled phenomena, such as the interplay between facial expression and dialogue in affective tutorial interaction.

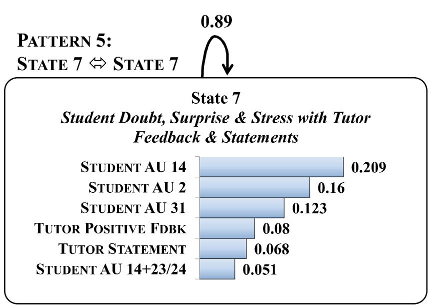
The facial expression and dialogue data described in Section 3 were merged into sequences of observations needed to build the HMM. Each observation consisted of a facial expression (denoted as facial action units (AUs) [7]), dialogue act or both. The Baum-Welch algorithm with log-likelihood measure was used for model training. Ten random initializations were performed to reduce convergence to local maxima. A hyperparameter optimization outer loop produced candidate HMMs across a range from three to twenty-two hidden states. Average log likelihood was computed across candidate HMMs for each number of hidden states. The models with best average log-likelihood had ten hidden states, and the best-fit model had the highest log-likelihood among these.

With the model in hand, the Viterbi algorithm was applied to map the most probable hidden state to each observation. Exhaustive search to length five across each session's hidden state sequences revealed five frequently recurring sequences (or "patterns") of affective tutorial interaction, shown in Figure 2. Each pattern occurred at a relative frequency greater than 0.05 across multiple sessions. Seven (of ten) hidden states comprised the patterns.

In order to examine the persistence of the five frequently-occurring patterns of affective tutorial interaction, average sequence lengths were calculated for each session (shown in Figure 2). There are subtle differences between relative frequency as a measure of prevalence and average sequence length as a measure of persistence. When the measures agreed (as was often the case), they showed prevalence and persistence of specific patterns of affective tutorial interaction within a particular session. When the measures differed, a persistent pattern recurred in long, but rare, sub-sequences or a prevalent pattern recurred in short sub-sequences.

The average sequence lengths shown in Figure 2 indicate notable differences in affective tutorial interaction within sessions. Thus, it may be possible to group sessions that have similar quantitative profiles. For instance, sessions 6 and 7 both have persistent sequences of PATTERN 2 and PATTERN 4, indicative of persistent student confusion with tutor statements and conversational dialogue during those sessions. Likewise, PATTERN 1 models tutor lecturing and instruction with occasional student participation and student affective states, PATTERN 3 is dominated by student facial displays (mostly surprise and frustration), and PATTERN 5 is largely composed of doubt, surprise, and stress with occasional tutor feedback and statements. In this way, quantitative application of HMMs provides insight into profiles of affective tutorial interaction across tutoring sessions.





Average sequence length of HMM patterns

Session	P1	P2	P3	P4	P5	Other
1	**5.6**	0	**7.8**	**9**	1.5	1.3
2	3.8	**5.8**	**7.6**	3.5	3.5	1.3
3	**5.9**	2.7	4	2	**12**	1.2
4	3.2	5.6	2.8	4.5	**23**	1.4
5	3.6	3	**6.6**	2.7	4.8	1.5
6	1.8	**9.7**	2	**9.3**	0	1.6
7	5	**8.3**	4	**5.4**	5	1.5

Fig. 2. Five patterns (i.e. frequently recurring sequences of hidden states) of affective tutorial interaction discovered from the best-fit hidden Markov model. Transition probabilities ≥ 0.5 are displayed. Emissions probabilities ≥ 0.05 are shown.

5 Conclusion and Future Work

The descriptive HMM learned from facial expression and task-oriented tutorial dialogue revealed five frequently-occurring patterns of affective tutorial interaction. Each pattern modeled distinct and interpretable segments of the tutoring sessions. A closer inspection of hidden state sequences as they occurred within sessions showed notable differences between sessions.

While this approach toward semi-automated understanding of affective tutorial interaction was successful, there are two primary limitations that highlight important directions for future work. First, manual FACS coding requires substantial manual labor, although this may become irrelevant when sufficient reliability is achieved in automated facial expression recognition. Second, the small sample size was a limiting factor, but using this approach across more tutoring sessions may identify statistical relationships involving discovered patterns of affective tutorial interaction. The quantitative distinctions in prevalence and persistence of discovered patterns of affective tutorial interaction may highlight individual or group-wise differences, leading to correlational analyses of HMM patterns and tutorial outcomes, such as self-efficacy and learning gains.

Further studies investigating the application of machine learning techniques are merited to advance the state of semi-automated affect understanding. Leveraging novel, semi-automated techniques may enable us to better understand affect during learning and contribute to efforts to integrate affect in ITSs.

Acknowledgements. This work is supported in part by the North Carolina State University Department of Computer Science along with the National Science Foundation through Grant DRL-1007962 and the STARS Alliance Grant CNS-1042468. 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. D'Mello, S.K., Calvo, R.A.: Significant Accomplishments, New Challenges, and New Perspectives. In: Calvo, R.A., D'Mello, S.K. (eds.) New Perspectives on Affect and Learning Technologies, pp. 255–271. Springer, New York (2011)
2. D'Mello, S.K., Lehman, B., Person, N.: Monitoring Affect States During Effortful Problem Solving Activities. International Journal of Artificial Intelligence in Education 20 (2010)
3. D'Mello, S.K., Graesser, A.: Modeling Cognitive-Affective Dynamics with Hidden Markov Models. In: Proceedings of the 32nd Annual Cognitive Science Society, pp. 2721–2726 (2010)
4. Johns, J., Woolf, B.: A Dynamic Mixture Model to Detect Student Motivation and Proficiency. In: Proceedings of the 21st National Conference on Artificial Intelligence, pp. 163–168 (2006)

5. Grafsgaard, J.F., Boyer, K.E., Lester, J.C.: Predicting Facial Indicators of Confusion with Hidden Markov Models. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011, Part I. LNCS, vol. 6974, pp. 97–106. Springer, Heidelberg (2011)
6. Boyer, K.E., Phillips, R., Ingram, A., Ha, E.Y., Wallis, M., Vouk, M., Lester, J.: Characterizing the Effectiveness of Tutorial Dialogue with Hidden Markov Models. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010. LNCS, vol. 6094, pp. 55–64. Springer, Heidelberg (2010)
7. Ekman, P., Friesen, W.V., Hager, J.C.: Facial Action Coding System. A Human Face, Salt Lake City, USA (2002)
8. Rabiner, L.R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE 77, 257–286 (1989)