Automated Posture Analysis for detecting Learner's Interest Level

Selene Mota and Rosalind W. Picard
MIT Media Laboratory
20 Ames Street
Cambridge, MA 02139, USA
{atenea, picard}@media.mit.edu

Abstract

This paper presents a system for recognizing naturally occurring postures and associated affective states related to a child's interest level while performing a learning task on a computer. Postures are gathered using two matrices of pressure sensors mounted on the seat and back of a chair. Subsequently, posture features are extracted using a mixture of four gaussians, and input to a 3-layer feed-forward neural network. The neural network classifies nine postures in real time and achieves an overall accuracy of 87.6% when tested with postures coming from new subjects. A set of independent Hidden Markov Models (HMMs) is used to analyze temporal patterns among these posture sequences in order to determine three categories related to a child's level of interest, as rated by human observers. The system reaches an overall performance of 82.3% with posture sequences coming from known subjects and 76.5% with unknown subjects.

1. Introduction

The aim of this research is to extract, process and model sequences of natural occurring postures for the purpose of interpreting affective states that occur during natural learning situations. The purpose of automating such recognition is twofold: first, to inform theoretical understanding of human behavior in learning situations; second, to enable the development of computerized learning companions [14] that could provide effective personalized assistance to children engaged in learning explorations.

Our emphasis has been on naturally occurring postures, and on interpreting these in ways that are semantically meaningful to human observers. It is crucial the system be capable of dealing with the unconstrained nature of real data, and be able to "discover" what children do naturally.

Our main hypothesis in this work is that certain kinds of affective information show up in the posture channel during natural learning situations, and that systems can be designed to automate detection of such information. To our knowledge, this hypothesis has never been systematically examined prior to our efforts, and this research is the first to support this hypothesis.

Several challenges exist in this work. First, researchers in the field of non-verbal behavior have not established a generalized criterion about how to classify postures [4,15,17]. As a consequence, there is not a well-established agreement about what are "basic posture units".

Second, there is not a clearly articulated association between postures and their interpretation. Ekman [8] proposed that body movements only carry information about the intensity of the emotion that is being experienced. In contrast, Bull [3] presented results showing that both body movements and positions transmit information about four distinctive emotions and attitudes. In the field of learning, some studies have presented empirical evidence about the correlation between postures and the student's level of engagement in a lesson [10,11]. On non-verbal behavior and discourse, Cassell J. et al. [5] have looked at the correlation of the dynamics of postural shifts and the introduction of a new topic; they found that postural shifts might signal boundaries of information units.

Third, there are a wide range of methods for detecting postures automatically. Many researchers have used cameras as input devices. However, using vision, the posture recognition is complicated by variations in the lighting or background conditions, camera or subject positions, and subject appearance. Consequently, there have been attempts to use other sensors such as switches, accelerometers, or pressure sensors mounted on a chair [9,24,25].

In this work, we consider only posture recognition of a person seated in a chair. We further discriminate two kinds of recognition: recognition of a static posture position, and recognition of a sequence of postural behaviors. Although we do not anticipate that the static position (e.g. leaning forward) can be reliably associated with an affective state (e.g. interest) we do hypothesize, and find evidence to support, that sequences of postural behaviors (e.g. repeatedly leaning forward, even if punctuated by occasional movements backward) can be used to predict significant information related to affective state. Thus, this work focuses both on static posture analysis and on analyzing patterns of posture over time.

The closest work to ours is that of Tan et al. [25, 26], who have used the same chair-based system that we use. Their effort has focused on recognition of static postures made by adults who intentionally position themselves into one of less than ten postures as requested by the experimenter. In contrast, we gather data of naturally occurring posture actions, and analyze not only static posture, but also postural sequences, with the aim of discovering affective interpretations associated with postural behaviors. Thus, we propose, build, and evaluate a system that integrates the recognition of postures together with their human interpretation.

The rest of this paper presents this work in four parts: Data collection (Sec 2); Selection of the set of posture action units and set of affective states (Sec 3); Feature extraction and recognition of posture units in real time (Sec 4); Temporal modeling and recognition of interest from posture sequences (Sec 5).

2. Data collection

2.1. System Hardware

Postures are recognized using two matrices of pressure sensors made by Tekscan [27]. One matrix is positioned on the seat-pan of a chair; the other is placed on the backrest. Each matrix is 0.10 millimeters thick and consists of a 42-by-48 array of sensing pressure units distributed over an area of 41 x 47 centimeters. A pressure unit is a variable resistor, and the normal force applied to its superficial area determines its resistance. This resistance is transformed to an 8-bit pressure reading, which can be interpreted as an 8-bit grayscale value and visualized as a grayscale image. In our experiments, we use for each sensing sheet a temporal resolution of 8 frames per second.

The two sensing sheets are placed on a SteelCase Leap chair. This chair was selected because it can be adjusted to a wide range of subjects' sizes (seat pan and back rest altitude & openness) and it has a firm, yet comfortable curved seat pan that provides basic ergonomic requirements of comfort.

2.2. Human experiment

Data collection for affective studies is a challenging task: The subject needs to be exposed to conditions that can elicit the emotional state in an authentic way; if we elicit affective states on demand, it is almost guaranteed not to bring out genuinely the required emotional state.

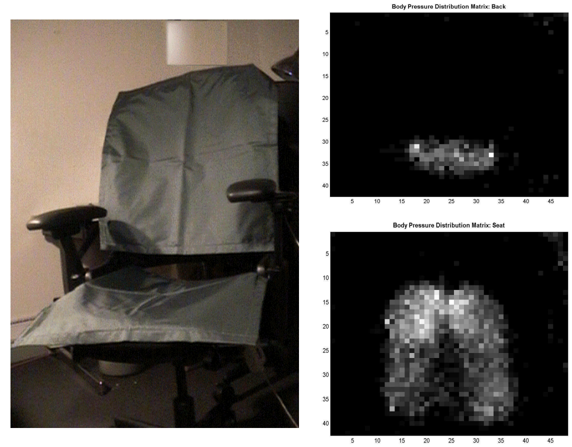

Figure 1. Left: Leap chair with pressure sensors. Right: Example pressure patterns for back (top image) and seat bottom (lower image.)

Affective states associated with interest and boredom were elicited through an experiment with 10 children (5 male and 5 female) ages 8 to 11 years, coming from relatively affluent areas of the state of Massachusetts in the USA and from a variety of cultural and economic backgrounds. Each child was asked to solve a constraint satisfaction game called *Fripples Place* [7] for approximately 20 minutes, while sitting in the chair lined with pressure sensors. The children were told that we wanted to know how fun, friendly, and interesting the game was. A parent was told fully about the study, and parental consent was obtained before proceeding. The experiment was conducted under the approval of the MIT Committee on the Use of Humans as Experimental Subjects.

The space where the experiment took place was a naturalistic setting designed to allow the subject to move freely and naturally. It was arranged with the sensor chair, one computer playing the game and recording the screen activity, having a 21" inch monitor, standard mouse and keyboard, as well as two Sony video-cameras: one capturing a side-view "posture video" and one the frontal view "face video" of the subject; and finally, an IBM Blue Eyes camera [12] capturing the face. We made the cameras unobtrusive to encourage more natural responses preserving as much as possible the original behavior.

The experiment procedure was as follows: First, the experimenter introduced herself and then, she performed a short interview with the subject. Subsequently, the experimenter showed the Fripples Place game to the participant and gave him general instructions about how to play it. The participant was then asked to play the game once and ask any questions about it. Lastly, the participant was instructed to play alone for around 20 minutes, until the experimenter returned.

Of the five channels of data collected (pressure patterns, posture video, face video, computer screen activity, and BlueEyes video) only the first four are used in this paper. The first two video streams and the computer screen activity were used to carry out two psychology studies: one for determining the set of posture action units and the other for establishing the set of affective states to be recognized by the system. The pressure data was the only channel used for the computer analysis in this paper.

3 Posture action units and affective states

3.1. Posture action unit selection

This work follows the philosophy proposed by Peter E. Bull [4] in his Body Scoring System. It uses movements or actions rather than positions of the body parts as the basic unit of analysis; hence it is possible to describe postures as a series of actions rather than as a series of positions, capturing the natural structure of body movement.

Two adult coders (graduate students) labeled postures using data collected by one of the video channels: the side-view posture videos. Two data sets were created. Each data set was formed using 100 ten-seconds-long segments randomly extracted from 10 children's posture videos. After coders iterated among different potential labels using the first data set, a final collection of nine posture actions was agreed upon: *sitting on the edge, leaning forward, leaning forward right, leaning forward left, sitting upright, leaning back, leaning back right, leaning back left and slumping back*. Subsequently, coders classified the second data set utilizing these posture categories along with 3 levels of confidence: low, medium and high. They obtained an overall agreement of 83 percent[1] using Cohen's Kappa formula [1].

The chair pressure patterns were synchronized with the video segments. Posture samples labeled with high and medium level of confidence were taken to compose the data sets for training and testing the static posture recognition system.

3.2. Affective state label selection

Given that we cannot directly observe the student's internal thoughts and emotions, nor can children in the age range of 8 and 11 years old reliably articulate their feelings, we chose to focus on labeling behaviors that communicate affect outwardly to an adult observer. This choice was made in consultation with Jerome Kagan at Harvard, an expert in experiments with children [13].

Table 1: Cohen's kappa between teachers	**%**
Teacher 1 and Teacher 2	78.00
Teacher 1 and Teacher 3	84.32
Teacher 2 and Teacher 3	73.39
Average of above	78.57

[1] According to Robson C. [22], Kappa in the range 0.4 to 0.6 is fair, between 0.6 and 0.75 is good and above 0.75 is excellent.

Choosing affective descriptions relevant to the learning experience is itself a challenge. We engaged in several iterations with sets of teachers in an effort to ascertain a set of labels that were meaningful and could be reliably inferred from the data. First, we made the decision to allow the teachers to look at frontal video, side video, and screen activity, recognizing that people are not used to looking at chair pressure patterns. Although this gave them access to more channels than we planned to give the computer (limiting it to chair pressure analysis, at least initially) we hoped this would lead to the most accurate and humanly meaningful affective labels. Eventually, we found that teachers could reliably label the affective states of *high, medium and low interest, taking a break*[2]*, bored and other*. Working separately and without being aware of the final purpose of the coding task, teachers obtained an average overall agreement of 78.57% (Table 1.)

In work below, we did not use data classified as *bored* or *other* even though teachers identified them consistently. The *bored* state was dropped since teachers only classified five episodes as bored, and this was not enough to develop separate training and test sets. The *other* state was eliminated because it was almost always associated with facial expressions and not postural behaviors.

All five channels of data were synchronized with the teacher's labels. As a result, we obtained approximately 200 minutes of labeled data, around 20 minutes per child.

4. Automated recognition of posture action units

4.1. Noise removal

We apply two methods for cleaning noise from the data. The first kind of noise was caused by deformations that the sheets suffer when they are placed on the seat-pan and backrest of the chair. This noise is removed by thresholding: finding the highest pressure value, multiplying it by 0.1, and eliminating those values or smaller from further analysis. The second method applied to the raw data is the morphological operation of erosion [23]. This operation is used to delimit the shape and boundaries of the body pressure distribution image as well as to reduce unwanted noise produced by the sensor

[2] Taking a break – where the subject shifts position between learning forward and back, sometimes with hands stretched above the head - is probably not an affective state, but this behavior appeared to be informative with respect to distinguishing postural activity related to boredom and interest.

itself. A square element with an anchor point of 3 (matrix 1-by-3) was used as a kernel.

4.2. Feature extraction: mixture of Gaussians

After noise removal, the sampled pressure points are modeled with a mixture of Gaussians. The pressure distribution data presents a geometrical structure that we can describe with four clusters in the 3-dimensional space. The number four (as a maximum) was chosen after testing with several configurations. Thus, a posture is compactly represented by the gaussian parameters: prior probability, mean and variance. These make up the vector of posture features.

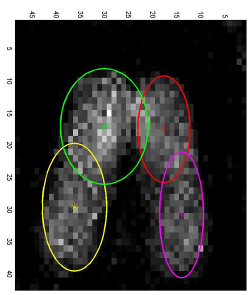

Figure 2. Seat pressure distribution matrix modeled with 4 gaussians. Each circle represents the parameters (mean and variance) of every gaussian.

The Gaussians are estimated using unsupervised clustering via the expectation-maximization algorithm [21]. Additionally, we constrain the algorithm to preserve the relative positions between the Gaussians: we implement this constraint via the termination criteria. With the usual Gaussian mixtures algorithm if the subject has just one leg on the chair, the four Gaussians will be distributed on that area, and then, the algorithm might not be able to discern that the parameters represent just one leg. In contrast, the modified algorithm can distinguish if just one leg is lying on the chair.

4.3. Static posture classification

After fitting the Gaussian mixture, the gaussian parameters (mean and variance) are used to feed a 3-layer feed-forward neural network [2] that classifies the static set of nine postures in real time.

The neural network input vector contains 56 elements coming from eight gaussians (four for each pressure matrix). Each gaussian has 7 parameters: 3 for the mean position, 3 for the covariance matrix diagonal, and one for the prior. The neural network architecture is structured as follow: the first layer comprises 56 neurons, the second 12, and the third 9. Their transfer functions are: a tan-sigmoid in the first layer; a log-sigmoid in the second layer; and a linear function in the third one. During its training phase, the neural network used the Bayesian Regularization Algorithm [16] along with the Minimum Square Error as the performance function for adjusting its weights and biases. This algorithm was implemented using the Matlab Neural Networks Toolbox [6]. Once the training phase was completed, the neural network was implemented in C++ language for real-time performance.

Posture samples coming from ten subjects constitute the static posture data set: 9634 samples from five subjects were used for training, whereas 7710 from the other five were used for testing. Table 2 summarizes the results obtained on the testing data.

Table 2: % Static Posture Classification	
Leaning Forward	96.68%
Leaning Forward Left	80.02%
Learning Forward Right	76.65%
Sitting Upright	93.21%
Leaning Back	90.91%
Leaning Back Left	79.86%
Leaning Back Right	89.43%
Sitting on the Edge of Seat	91.91%
Slumping back	90.12%
Total	**87.64%**

The results show the neural network classifies postures coming from new subjects with an overall accuracy of 87.64%. In specific, the network is capable to classify within a range of 90% to 97% postures belonging to the classes of sitting on the edge, sitting upright, leaning forward, leaning back, and slumping back. However, its classification range goes down, 77% to 90%, when the network is tested with postures from the other four classes: leaning forward right, leaning forward left, leaning back right, and leaning back left.

5. Recognizing interest from posture sequences

5.1 HMM modeling

A set of independent Hidden Markov Models (HMMs) [20] was developed for modeling posture sequences for the affective states identified by the teachers: high interest, medium interest, low interest and taking a break. Each HMM takes a sequence of postures obtained in the previous neural net layer as discrete inputs. Given an HMM that represents each affective state, the system computes the probability that the observed posture sequence was produced by that HMM. Hence, the observed posture sequence is classified according with the HMM model with the highest probability.

Each HMM (H) is determined through the set of posture symbols *V*, the sequence of observed postures O, the length of the observation sequence (T), the number of hidden states (N), their initial probabilities (pi), the hidden states transition matrix (A), and the emission probability matrix (B). To determine the HMM parameters, we used the observed posture data not only for parameter estimation but also for model selection. We estimated the HMM model parameters using the Baum-Welch algorithm [20], implemented using Kevin Murphy's Matlab HMM Toolbox [18].

K-fold cross-validation [19] was used as the training method for determining the parameters T and N, choosing the parameters that gave the smallest generalization error. Specifically, the full set of children's posture sequences was randomly divided into k=10 sub-groups of approximately equal size. The model parameters were estimated 10 times, each time leaving out one of the subgroups from training, but using only the omitted subgroup to compute the generalization error. We use the log likelihood [20] as the evaluation function:

$$\log P(O_t, H_t) = \log P(H_1) + \sum_{t=1}^{T} \log P(O_t \mid H_t) + \sum_{t=1}^{T} \log P(H_t \mid H_{t-1})$$

The HMM's were able to correctly classify sequences that range from 24 to 88 posture observations; hence, the models started to differentiate affective properties of posture sequences after accumulating samples for at least 3 seconds. We found the optimal HMM parameter arrangement was T=64 with N=9 for high interest, N=11 for low interest, and N=11 for the taking a break models. For models of medium interest, although we tried several combinations of N and T, none of them had good performance. The sequences coming from the medium interest category were most of the time confused between high interest and low interest. Thus, we decided to proceed using only the sequences coming from the classes of high interest, low interest, and taking a break.

5.2 Evaluation and Results

The system was evaluated two ways: First, the HMMs were tested using k-fold cross-validation with k=8. Posture data coming from 8 subjects was randomly divided into 8 groups, with each group having examples from multiple subjects. Then, each HMM was trained using data from 7 groups and reserving one group for testing. This was repeated for all the 8 groups forming the data set. Second, the HMM classifiers were tested using posture sequences coming from two subjects that were not in the training set (new subjects). In both cases the classifications of the posture sequences were chosen to maximize the log-likelihood probability of each posture sequence given the HMM model.

The results of these two evaluations are shown in Tables 3 and 4. An average accuracy of 82% is attained for the three affect-related categories given the 8-fold cross validation. For the harder evaluation, training on eight and testing on two randomly selected unseen subjects, the overall recognition accuracy dropped 5.8% to an average of 76.5%. Most of the drop appears to be because of a decline in recognition of "taking a break." Thus, we see in both tests, despite a relatively small number of subjects to train on, that there is significantly greater than random recognition of postural behaviors related to affective state.

Table 3: HMM Results: 8 subjects, k-fold cross validation

Data Set	High Interest	Low Interest	Taking a Break	%Recognition
High Interest	76	6	7	85.39%
Low Interest	22	82	6	74.55%
Taking a Break	8	4	79	86.81%
			Total	**82.25%**

Table 4: HMM Results: 2 new subjects

Data Set	High Interest	Low Interest	Taking a Break	%Recognition
High Interest	10	1	1	83.33%
Low Interest	5	18	3	69.23%
Taking a Break	1	2	10	76.92%
			Total	**76.49%**

6. Conclusions and future work

This paper has hypothesized and found evidence to support a relationship between patterns of postural behaviors and affective states associated with interest while a child is working on a learning task via computer. This work has never assumed that static postures reveal what a student is feeling inside; rather, observed patterns in the dynamics of the student's postures were found to disclose significant information related to the affective states of high interest, low interest and the related behavior of taking a break. The latter behavior, which can co-occur with several affective states occurring in learning, appears to be an important component that increases in frequency preceding boredom.

Nine postures in our data were found to be reliably recognized by humans, and an automated pattern recognition system was built to detect these. The system achieves an average accuracy of 87.6% when tested with data not included in the training set. This result is significant given that the data set contained natural occurring postures gathered with children in real learning-computer interaction, which we believe makes the problem harder than using an artificial data set where the subjects were asked to make explicit postures. This posture recognition system runs in real-time, and it has been proved to work in a user-independent way. It is currently trained on children and not on adults, but potentially the same algorithms could be used to re-train the system for any population of interest.

Also, posture sequences were identified by teachers as reliably representing certain states related to interest and boredom. These states were modeled using HMMs, and achieved an overall accuracy of 87% when tested with new posture sequences coming from students that were included in the training set. The system reached an overall accuracy of 76% when tested with posture sequences coming from two subjects that were not included at all in the training set.

Thus, in contexts where children are learning using computers, this system could provide substantial information about whether the computer is engaging the child and whether the frequency of taking a break patterns is increasing. We expect that these two states may be especially relevant for determining when not to interrupt, or when it is likely that the child might welcome an interruption, perhaps for assistance or encouragement.

As future work, we expect that the combination of the recognition of dynamical posture patterns along with other modalities (face, computer task behavior, and possible conversational input) will further disambiguate the user's state, and improve the ability of the computer to respond in a way that facilitates a productive and enjoyable learning experience.

7. Acknowledgments

We thank Deb Roy and Justine Cassell for their advice, Stefan Agamanolis for assisting us with ISIS for indexing videos, Joel Stanfield from Steelcase for providing the Tekscan sensor system, and Hong Tan and Lynne Slivovsky for helping us to set up the chair hardware and sharing with us their previous findings.

8. References

[1] Bakeman, R., Gottman, J. (1986). *Observing Interaction: An Introduction to Sequential Analysis*, Cambridge University Press.

[2] Bishop, Christopher M. (1995). *Neural networks for pattern recognition*, Oxford: Clarendon Press; N. Y.: Oxford University Press.

[3] Bull P.E. (1983). *Body movement and interpersonal communication*, Chichester: John Wiley & Sons Ltd.

[4] Bull E. P. (1987) Posture and Gesture (16), Pergamon Press.

[5] Cassell, J., Nakano, Y., et al., *Non-verbal Cues for Discourse Structure*, Association for Computational Linguistics Joint EAC-'01 ACL Conference.

[6] Demuth H., Beale M., (1992 - 2001). Matlab Neural Networks Toolbox v. 4.0. MathWorks Inc: http:\\www.mathworks.com

[7] Edmark, Fripple Place. http://www.riverdeep.net/edconnect/softwareactivities/critical_thinking/fripple_place.html

[8] Ekman, P. (1965). *The differential communication of affect by head and body cues*. Journal of Personality and Social Psychology, 2,726-735

[9] Evreinov G., Agranovski A., et al. (1999). *PadGraph*, Proceedings of HCI International '99,vol. 2, pp. 985-989, Munich, Germany.

[10] Goldin-Meadow, S., D. Wein, et al. (1992). *Assessing Knowledge Through Gesture: Using Children's Hands to Read Their Minds*. Cognition and Instruction 9 (3): 201-219.

[11] Goldin-Meadow, S., M. W. Alibali, et al. (1993). *Transitions in Concept Acquisition: Using the Hands to Read the Mind*. Psychological Rev 100 (2): 279-297.

[12] IBM Blue Eyes Camera: http://www.almaden.ibm.com/cs/blueeyes

[13] Kagan, J. (1994). *The Nature of the Child: Tenth anniversary edition*, New York: Basic Books.

[14] Kapoor, A., Mota, S., & Picard, R. W.(2001) *Towards a Learning Companion that Recognizes Affect*, AAAI Fall 2001, MA, USA.

[15] LaFrance, M. (1982). *Posture Mirroring and Rapport. Interaction Rhythms: Periodicity in Communicative Behavior*. M. Davis. New York, Human Sciences Press, Inc.**:** 279-298.

[16] MacKay, D. J. C., (1992). *Bayesian interpolation*, Neural Computation 3 (4): 415-447.

[17] Mehrabian A., Friar J. T. (1969). *Encoding of attitude by a seated communicator via posture and position cues*, Journal of Consulting and Clinical Psychology, vol. 5.

[18] Murphy, K., (2002). *Hidden Markov Model Toolbox*, http://www.cs.berkeley.edu/~murphyk/Bayes/hmm.html

[19] Plutowski, M., Sakata, S., White, H. (1994). *Cross-validation estimates IMSE*. Advances in Neural Information Processing Systems 6: 391-398, Morgan Kaufman.

[20] Rabiner L. R. (1989). *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, IEEE 77 (2): 257-285.

[21] Redner R A., Walker H. F., (1984). *Mixture densities, maximum likelihood and the EM algorithm*, SIAM Review 26 (2): 1995-239.

[22] Robson C. (1993). *Real Word Research: A Resource for Social Scientist and Practitioner Researchers*, Oxford: Blackwell.

[23] Serra, J. (1982). *Image Analysis and Mathematical Morphology*, Academic Press, London.

[24] Smith J. (2000). *GrandChair: Conversational Collection of Family Stories*, M.S. thesis, MIT, Media Laboratory, Cambridge, MA, USA.

[25] Tan H. Z., Ifung Lu , Pentland A. (1997). *The chair as a novel haptic user interface*, In Proceedings of the Workshop on Perceptual User Interfaces, Banff, Alberta, Canada.

[26] Tan H. Z., Slivovsky L. A., and Pentland A., (2001). *A sensing Chair Using Pressure Distribution Sensors*, IEEE/ASME Transactions on Mechatronics 3 (6).

[27] Tekscan (1997). *Tekscan Body Pressure Measurement System User's Manual*. Tekscan Inc., South Boston, MA, USA.