A Bayesian Network Approach to Interpret Affective States from Human Gestures: An Application to Affective Tutoring

Abdul Rehman Abbasi

Asian Institute of Technology,
P.O.Box No.4, Klong Luang, Pathumthani 12120, Thailand
AbdulRehman.Abbasi @ait.ac.th

Abstract. Vision-based human affect analyses is an active research area for development of intelligent and interactive affective computing systems. These systems would be more effective, if they reliably interpret the affective signals. Current research is mainly focused on improving system's accuracy and efficiency, to detect and recognize visually observable signals such as face and body gestures. In reality, the relationship between these signals and their subjective interpretation by humans is quite noisy. In this paper, we propose an approach that relates few human gestures to the most probable affective states in a given situational context. We utilize Bayesian network model to interpret human affect using the knowledge base, developed on the basis of our pilot study. This study was carried out to understand the affective response of students attending a class lecture. We have extracted some meaningful gestures out of that response, and propose real time detection and interpretation of these gestures to benefit real world applications such as affective tutoring.

Key words: Affective tutoring, Bayesian networks, gesture analyses, knowledge-base, situational affect interpretation.

1 Introduction

1.1 Motivation and Problem Statement

Imagine a scenario: you are attempting to learn a complex tutorial through an affective tutoring system. During the proceedings, you start scratching your head. Realizing your anxiety, the intelligent tutoring system decides to give you some useful information about the current topic that resolves your anxiety and you get back to the proceedings again. The described scenario reflects the ability of the system to interpret human gestures, and use them for useful feedback to its user. Promising applications such as this and others enumerated by [1], are source of inspiration for growing interest in researching naturally-coupled human to computer interactive systems. But, these systems will only perform well and up to their users' expectations, when they would *reliably interpret* their affective state.

At present, vision-based affect analyses research is focused on improving the efficiency of affect recognition methods and hardly any attempt could be found that evaluates the reliability in recognition of affective states from visual signals. Subjective reporting is a convenient way to validate human affect while mapping gestures to affective states, though not the most convincing way, as it might be influenced by biases and expressivity, but designing affective systems for humans need their involvement in the process, so one should not overlook this aspect during analyses.

In fact, the mapping between visible gestures and their interpretation itself is quite noisy. The noisiness in mapping may be caused by inconsistency in how different people react to a situation. This inconsistency might be due to personality difference, social and situational context and needs to be formalized by some approach to interpret affective states reliably. Considering the *situational context* might be helpful to address this problem.

From our pilot study [2], to understand human affect under a certain situation, we found few consistencies in certain mappings and in this research, we aim to develop and evaluate such an approach which maps gestures to probable affective states based on the proposition that probability of mapping is a function of the situation under which these gestures were observed.

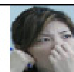
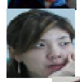
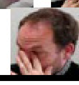
1.2 Overview of Existing Research

Do intelligent systems really need affective understanding of their users? Emotion is a part of human intelligence [3]. Minsky [4] says that "the question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without emotions." As human, we expect a lot from our intelligent counterparts and therefore we are quite justified to expect that intelligent systems should realize our feelings and adapt accordingly.

During communications, we try to guess intentions and feelings of others, from their behavior or body language. This behavior or body language is the way they react to a situation or an event, and it reflects their affective state. An affective state can be considered as instance when a person feels some emotion such as "Sadness" or have certain mental state such as "Recalling". As such no precise and generally agreed definition of emotion exists. Recently, [5] describes emotional states as not different from the processes such as thinking.

Humans' intention, state of mind and emotion can be readily estimated by variety of channels. These channels are verbal and non-verbal. In nonverbal human to human communications, face and gestures are commonly used sentic modulations [1]. Basically, sentic modulations are observable clues for affect mapping and these include facial expressions, vocal intonations, gestures, postures and other bodily changes. Darwin [6] emphasized the importance of facial expressions as powerful and immediate means of communications. Mehrabian's work [7] also suggests that human facial expressions based signals convey about 55 percent of the actual message. Inspired by these ideas, many researchers have made valuable contributions by automatically extracting and analyzing such signals. Comprehensive reviews related to such efforts can be found in [8–13].

Table 1. Selected human gestures for the proposed model

S.No.	Image	Action	Interpretation by Subjects
1	 	Head Scratching, Nose Scratching	Recalling
2	 	Chin Resting Lips Touching	Thinking
3		Rubbing Eyes	Tiredness

These automated systems are designed to extract emotions out of the available visual responses. The idea is to map facial expressions and gestures to certain emotional categories. Some of them objectively classify facial actions using Ekman's Facial Action Coding System [14]. In fact, the majority of the work involves mapping *stereotypical posed expressions,* lacking spontaneity. Strong evidences of the universality of certain facial expressions based on extensive research studies by Ekman [15], supports the presence of few prototypical facial expressions. But these expressions are subject to display rules, governed by societal norms and cultural backgrounds [16].

Sebe et al. [17] also reported their work on extracting spontaneous facial expressions. However they observed that: (a) it was quite difficult to get a wide range of emotions for all of the subjects, and (b) the facial expressions corresponding to internal emotions are often misleading.

In a recent pilot study [2], we found some observable gestures, shown in Table 1, that were probabilistically related to particular affective states in the context of a classroom lecture. This lead us to propose, towards enhanced human-computer interaction in intelligent tutoring systems, a framework for context-specific probabilistic interpretation of gestures and an empirical evaluation of that framework in a real-world classroom tutoring situation.

1.3 Objectives of Research

Whenever there exists uncertainty, we turn to probability as a means to cope with uncertain situations. Interpreting affective states from gestures involves uncertainty due to reasons described above, so we need to tune our system accordingly. Bayesian networks are found to be quite useful to predict with uncertain clues and incomplete information [18]. These are intuitive in nature and provide a relationship between cause and effect, linked by the probability. With a probabilistic inference of a particular gesture by a human in a given situation, reliability of affect recognition system might increase. For the current research, we propose to use Bayesian network and as proof of concept, we will focus on the small knowledge base as illustrated in Table 1.

Many researchers have successfully employed probabilistic methods for affect recognition and emotion classification but as per our knowledge from the available literature, our proposed approach is the first ever effort towards *probabilistic interpretation* of human affect from body gestures for a *specific scenario*. State of the art work on gesture recognition [19], may be utilized for gesture detection, prior to its probabilistic interpretation in the situational context.

The results obtained by this research may be extended to other situational affect recognition applications. It may provide a feasible solution for reliable affect recognition in a world of uncertain human behavior.

The rest of paper is organized as follows: Section 2 describes the methodology for the proposed system. Section 3 highlights the system implementation. Section 4 concludes the discussion with expected contribution.

2 Methodology

Conceptually, our proposed model is based on utilizing the small domain knowledge that we formed from our pilot study [2]. It consists of three basic modules (Refer Fig. 1(a)). These modules are Action Detection, Inference and Decision Action module. The Action Detection module estimates the probability of the presence of a set of particular human actions (mentioned in Table 1) which are interpreted by the Inference module. Domain knowledge will help in relating human actions to probable affective states. Finally, the Decision Action module will provide user interface showing the recognized state. Each module is described in detail in sections to follow.

2.1 Action Detection Module

As the human gestures (actions) that we desire to detect involve hand movement relative to face, we propose to use skin color-based segmentation. Skin color based information retrieval methods are quite prevalent and are oftenly used in human-robot interaction applications [20, 21]. The detection of the relative position of the hand from the face will identify the presence of a gesture as described in Table 1. The feature vector retrieved from the segmented image will be used as input to the classifier. Template feature vector for the target gestures will be used to classify the input vector accordingly.

2.2 Domain Knowledge

The domain knowledge provides mappings between detected actions and affective states using a Bayesian network. Graphical models interfaced to data via probability theory, such as Bayesian networks provide a simplified way to solve problems that involve uncertainty and complexity [22]. They provide good results even with insufficient data. Bayesian inference uses a numerical estimate of the degree of belief in a hypothesis before evidence has been observed and calculates a numerical estimate of the degree of belief in the hypothesis after

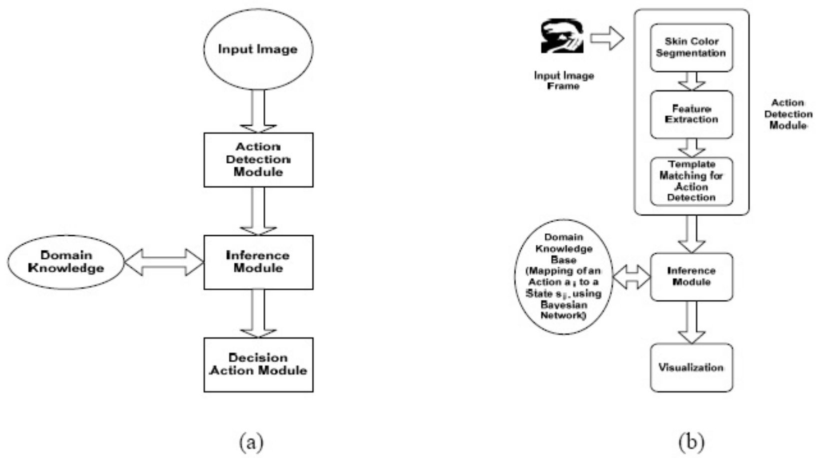


Fig. 1. (a). Conceptual model of proposed system, (b). Details of proposed system

evidence has been observed [23]. Bayesian inference usually relies on degree of belief, or subjective probabilities, in the induction process and does not necessarily claim to provide an objective method of induction. Bayesian probability is an interpretation of probability suggested by Bayesian theory, which holds that the concept of probability can be defined as the degree to which a person believes a proposition. It suggests that Bayes' theorem can be used as a rule to infer or update the degree of belief when new information is presented.

At present, Bayesian networks are widely used in artificial intelligence applications. These include medical diagnosis, image understanding, speech recognition, multi-sensor fusion and environmental modeling. Bayesian networks are quite attractive for modeling emotion and personality especially in embodied animated agents [24]. The reason is that they can cater nondeterministic phenomenon such as human emotional response under various situations. They can predict the likelihood of an affective state given a particular human action. Researchers do favor the usage of high level reasoning for best inference of emotions [25].

In our case, we have some probable affective states that are probabilistically related to some given actions. The value of probability will determine this mapping. We infer the outcome using Bayesian network model for possible states (Refer to Fig.2).

2.3 Inference Module

In fact, Bayesian network shows a relationship between cause and effect linked by the probability. For the retrieved data from our pilot study [2], we found the

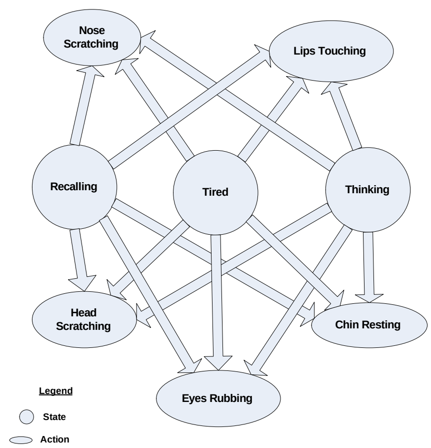


Fig. 2. Cause-Effect relationship of human gestures and affective states

actions (effects) and the affective states (causes) as shown in Table 1 and Fig.2. In the network of Fig.2, inference is straight forward application of Bayes' rule.

In equation (1), we define posterior probability $P(s_j|a_i)$ of an affective state s_j given a detected action a_i, using Bayes' rule:

$$P(s_j|a_i) = P(a_i|s_j).P(s_j)/P(a_i) \quad (1)$$

Where $P(a_i|s_j)$ is the likelihood of action a_i, given the state s_j while $P(a_i)$ and $P(s_j)$ are the prior probabilities for action a_i and state s_j respectively. These variables are interpreted as in relation (2) and (3):

$$a_i \epsilon [Head\ Scratching, ..., \ Lips\ Touching] \quad (2)$$

$$s_j \epsilon [Recalling, Thinking, Tired] \quad (3)$$

2.4 Decision Action Module

The Decision Action module provides an interface between the user and the system. In our prototype, it will be a graphical user interface showing probable affective state of the person (Refer to Fig.3). It provides the interpretation of the affective state on the basis of inference made by the inference module.

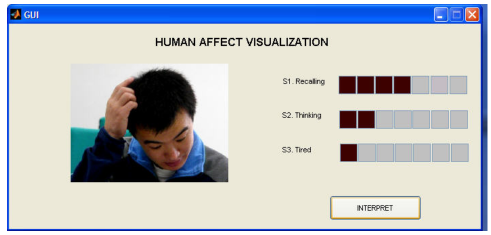


Fig. 3. Visualization by Decision Action Module

3 System Implementation

The proposed system will be trained using the data available from our pilot study along with other positive and negative training examples. Testing will be performed by real time deployment of the system. Input parameters to be used are human actions, as detected by Action Detection Module. Domain knowledge will describe the subjective interpretation in the specific scenario.

4 Contribution

Human affective state recognition is an active research area for a variety of reasons. These include the need of natural human-to-machine communication for applications such as human-robot interaction in a variety of domestic and industrial applications. Other applications that may benefit from these analyses may be categorized as educational such as affective tutoring. Affective applications will be less effective if they do not consider the uncertainty of the relationship between human gestures and underlying affect. With our proposed research, reliable gesture interpretation might be useful to produce effective systems that use these interpretations for enhanced man-machine interaction.

Acknowledgments. I am deeply indebted to my Program Committee including Dr. Nitin V. Afzulpurkar (Chair), Dr. Matthew N. Dailey (Co-Chair), Dr. Manukid Parnichkun (Member) and Prof. Erik L.J. Bohez (Member). Due acknowledgements for the participants of the experiments, conducted at National Institute of Informatics, Tokyo, Japan. Special thanks to Dr. Takeaki Uno (Assoc. Prof. at NII) for his generous support and cooperation.

References

1. Picard, R.W.: Affective Computing. MIT Press (2000)
2. Abbasi, A.R., Uno, T., Dailey, M.N., Afzulpurkar, N.V.: Towards Knowledge-Based Affective Interaction: Situational Interpretation of Affect. To Appear in Proc. 2nd Int. Conf. on Affective Computing and Intelligent Interaction, Lisbon, Portugal (2007)

3. Salovey, Mayer, J.D.: Emotional intelligence. Imagination, Cognition, and Personality **9**(3) (1990) 185-211
4. Minsky, M.: The Society of Mind. Simon & Schuster, New York (1985)
5. Minsky, M.: The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. Simon and Schuster, New York (2006)
6. Darwin, C.: The Expression of the Emotions in Man and Animals. Oxford University Press (1872/1998)
7. Mehrabian, A.: Communication without Words. Psychology Today **2**(4) (1968) 53-56
8. Tian, Y., Kanade, T., Cohn, J.F.: Facial Expression Analysis. In: Li, S.Z., Jain, A.K. (eds.): Handbook of Face Recognition, Springer (2005)
9. Pantic, M., Sebe, N., Cohn, J.F., Huang, T.: Affective Multimodal Human-Computer Interaction. In: Proc. ACM. Int. Conf. Multimedia. (2005) 669–676
10. Pantic, M., Rothkrantz, L.J.M.: Towards an Affect-Sensitive Multimodal Human-Computer Interaction. Proc. IEEE **91**(9) (2003)
11. Fasel, B., Luettin, J.: Automatic Facial Expression Analysis: A Survey. Pattern Recognition **36**(1) (2003) 259–275
12. Cowie, R., Cowie, E.D., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion Recognition in Human-Computer Interaction. IEEE Signal Processing Magazine **18** (2001) 32-80
13. Pantic, M., Rothkrantz, L.J.M.: Automatic Analysis of Facial Expressions: The State of the Art. IEEE PAMI, **22**(12) (2000) 1424–1445
14. Ekman, P., Friesen, W.: Facial Action Coding System (FACS) Manual. Consulting Psychologists Press (1978)
15. Ekman, P.: Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique, Psychological Bulletin **115** (1994) 268-287
16. Matsumoto, D.: Ethnic Differences in Affect Intensity, Emotion Judgments, Display rule Attitudes, and Self-Reported Emotional Expression in an American Sample. Motivation and Emotion **17**(2)(1993) 107-123
17. Sebe, N., Lew, M.S., Sun, Y., Cohen, I., Gevers, T., Huang, T.S.: Authentic Facial Expression Analysis, Image and Vision Computing, In Press, available online 9 October, (2006)
18. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, California (1988)
19. Mitra, S., Acharya, T.: Gesture Recognition: A Survey. IEEE SMC-Applications and Reviews, **37**(3) (2007) 311-324
20. Kim, K.K., Kwak, K.C., Chi, S.Y.: Gesture Analysis for Human-Robot Interaction. Proc. 8th Int. Conf. Advanced Communication Technology (ICACT), Korea (2006) 1824-1827
21. Rehg, J.M., Kanade, T.: Visual Tracking of High DOF Articulated Structures: an Application to Human Hand Tracking. Proc. 3rd European Conf. on Computer Vision, LNCS **801** (1994) 35-46
22. Cowell, R.: Introduction to Inference for Bayesian Networks. In: Jordan, M.I.(ed.): Learning in Graphical Models. (1999) 9-26
23. `http://www.bookrags.com/wiki/Bayesian_inference`
24. Ball, G., Breese, J.: Emotion and Personality in a Conversational Agent. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.): Embodied Conversational Agents. (2000)
25. Jaimes, A., Sebe, N.: Multimodal Human Computer Interaction: A Survey. In Computer Vision in Human-Computer Interaction: Proc. ICCV Workshop on HCI. Beijing, China (2005)