

# Active Affective State Detection and User Assistance With Dynamic Bayesian Networks

Xiangyang Li, *Member, IEEE*, and Qiang Ji, *Senior Member, IEEE*

**Abstract**—With the rapid development of pervasive and ubiquitous computing applications, intelligent user-assistance systems face challenges of ambiguous, uncertain, and multimodal sensory observations, user's changing state, and various constraints on available resources and costs in making decisions. We introduce a new probabilistic framework based on the dynamic Bayesian networks (DBNs) to dynamically model and recognize user's affective states and to provide the appropriate assistance in order to keep user in a productive state. We incorporate an active sensing mechanism into the DBN framework to perform purposive and sufficing information integration in order to infer user's affective state and to provide correct assistance in a timely and efficient manner. Experiments involving both synthetic and real data demonstrate the feasibility of the proposed framework as well as the effectiveness of the proposed active sensing strategy.

**Index Terms**—Active fusion, affective state detection, Bayesian networks (BNs), user assistance.

## I. INTRODUCTION

THE FIELD of human-computer interaction has moved from studies focusing on friendly interfaces such as the graphical user interfaces, to those that seek to understand, explain, justify, and augment user actions, focusing on developing more powerful representations and inferential machinery [17]. One important application is to design and implement intelligent and automatic agents to assist users in their daily work and life for performance enhancement [9]. While progress is being made in user-modeling [3], augmented cognition [6], and adaptive user interfaces, the majority of existing systems continue to assume normative performance, and fail to adapt to user affects. A constellation of recent findings, from neuroscience, psychology, and cognitive science, suggests that emotion plays surprising critical roles in user's rational, functional, and intelligent behaviors [21]. In fact, the situations where affective considerations are most critical are precisely the types of situations where the consequences of the human-machine interaction failures are most severe. For example, every year, many people are injured in car accidents because drivers are in a dangerous state, including fatigue, nervousness, confusion, or being stressed. If we could detect these negative affective

states in a timely manner, and provide assistance in terms of appropriate alerts, we may prevent many accidents from happening. However, the development of such systems faces several great challenges including: 1) sensory observations of the user are often ambiguous, uncertain, and from sources of different modalities; 2) user's affective states are often dynamic and evolve over time; and 3) decisions about the user's need and the assistance must be rendered appropriately and in a timely and efficient manner under various constraints.

We introduce a probabilistic framework based on the dynamic Bayesian networks (DBNs) and information theory to simultaneously address these challenges. First, a generic hierarchical probabilistic framework for user modeling is introduced to model the visual sensory observations, and the profile and contextual information related to the user's affective state. Second, this framework dynamically evolves and grows to account for temporal change in sensory observations as a result of the change in user's affective state. Third, the proposed framework provides a mechanism that performs purposive and sufficing information integration in order to determine the user's status timely and efficiently. Specifically, our system first formulates an initial hypothesis about the user's current affective state and then actively selects the most informative sensory strategy in order to quickly and economically confirm or refute the hypothesized user state.

## II. USER MODELING AND ASSISTANCE

In [24], the authors summarize the challenges for user modeling including the need for large datasets, the need for labeled data, concept drift, and computational complexity. Jameson [14] reviews the user and student-modeling systems that manage uncertainty using statistical techniques, including Bayesian networks (BNs), Dempster-Shafer theory, and fuzzy logic. Recently, there has been a significant surge in using Bayesian networks (BNs) in user modeling, intelligent tutoring, and other related fields. In the sections to follow, we review efforts closely related to ours including plan recognition, user need inference, and user affective state assessment.

### A. Plan Recognition

Plans are descriptions of action patterns. They encode a user's intentions and desires. When building up the user model, there is an assumption that rational agents have a mental state. Pynadath and Wellman [22] present a Bayesian framework describing the context in which the plan was generated, the mental state and planning process of the agent, and the consequences of the agent's actions. The core part of their work is the "Belief-Pref-erence-Capability" model of the agent's mental states that is

Manuscript received June 9, 2003; March 30, 2004 and June 22, 2004. This work was supported in part by the U.S. Army Research Office under Grant DAAD19-01-1-0402. This paper was recommended by Guest Editor G. L. Foresti.

X. Li is with the Department of Industrial and Manufacturing Systems Engineering, University of Michigan, Dearborn, MI 48128 USA (e-mail: xylum@umich.edu).

Q. Ji is with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: qji@ecse.rpi.edu).

Digital Object Identifier 10.1109/TSMCA.2004.838454

used to model the user. The belief is the agent's knowledge of the state and the dynamics of the world. The preference is the agent's intentions impacting its behavior on the world. The capability is the self-model of the agent's available actions. The authors' method is demonstrated in traffic monitoring to predict the plan of a driver from observation of vehicle movements. This framework provides a general way to model plan recognition problem. However, it is not intended for active assistance.

### B. User-Need Inference and Assistance

Intelligent assistance systems need the ability to adaptively accommodate the user's specific need. In the READY system [1], the authors use DBNs in a dialog system to adjust the policy in providing instructions, based on the recognized time pressure and cognitive load of the user from observations including filled pauses, disfluencies, and errors. Adaptation is realized by a rule base that maps detected situations into actions. No active information collection is considered.

Microsoft is carrying out extensive research applying BNs to create intelligent software assistants. The Lumiere project is intended to help computer users with interactive interfaces [10]. DBN models are used to infer a user's goals and needs by taking into account the user's background, actions, and queries. Based on the utility theory of influence diagrams, the automated assistant provides customized help. This research addresses the issues in automatic assistance such as the timing and optimization of assistance. However, it does not focus on providing active information fusion that dynamically selects information channels.

DeepListener augments speech recognition in clarification dialogs by using DBN models [12]. The models infer about user intentions associated with utterances, such as affirmation, negation, reflection, and so on. Utility, in terms of costs and benefits, is calculated by assessing the cross product of situations and actions, through psychological experiments or the use of assessment tools. DeepListener, however, does not distinguish actions and sensory tests in utility calculation and relies heavily on immersive interaction. This may be effective sometimes, but dangerous other times because normally users are highly varied regarding operating skills and personalities, and thus more easily become resistant to such "overactive" interface. The Bayesian receptionist system [11] suffers from the same problems plaguing DeepListener. Its central goal decomposition hierarchy uses Bayesian models at increasingly detailed levels of analysis. At each level, the system applies a greedy value of information calculation based on entropy to select the next single piece of evidence. When the expected cost of evaluating observations exceeds the expected value, the value of information calculation terminates within the current level and the system moves to the next level of detail.

### C. Affective Computing

More and more HCI researchers are interested in users' emotional and mental aspects, since affective states are an important indication of the user's internal state, intention, and needs. Affective computing focuses on emotional intelligence recognition [20]. Human beings have abundant emotions, such as sadness, happiness, guilty, pride, shame, anxiety, fear, anger, etc. From the view of computational theory, affective-state assess-

ment uses pattern recognition and information retrieval technologies. These techniques include fuzzy rules [13], Bayesian learning [23], hidden Markov models (HMMs) [4], BNs, etc. Most of these research efforts focus on the low level mapping between sensory data and underlying emotions. We categorize them into two groups. The first group uses sensory measures as predictors and applies classification algorithms without the prior and context knowledge about these variables and the target affective states. In building pattern models and committing classification tasks, such algorithms lack the ability to handle uncertainty, complexity, and ambiguity involved in data. The second group, represented by BNs and HMM models, represents the prior knowledge and expertise in graphical network form. They maintain the balance between global and local representations and the built-in causal and uncertainty representation structure provides powerful capabilities in handling complex situations in practical systems.

Ball and Breese [2] use a BN to assess the user's affective state in terms of valence and arousal and the personality in terms of dominance and friendliness. The observable data are facial and speech information about the user. A DBN model is used to capture the temporal emotion-state structure for a simple emotional state assessment task. Conati [5] provides a DBN model for assessing students' emotion in educational games. The emotion states are modeled as consequence of how the current action and help fit with the student's goals and preferences. Some body expressions are also used as evidences.

### D. Fault Detection and Troubleshooting

Finally, another area of research that is related to the proposed active sensing is fault detection and troubleshooting. Fault diagnosis and troubleshooting are decision-theoretic processes that generate low-cost plans for identifying faults so that a device can be repaired efficiently. Heckerman *et al.* [8] apply an approach based on BNs to encode the possible faults and to identify an optimal repair action plan, achieved by evaluating the cost of various plans. After each action, the probabilities are updated and new potential plans are generated. Langseth and Jensen [16] extend the traditional troubleshooting framework to model non-perfect repair actions and questions. The efficiency, the expected cost of repair, and the value of information for actions and questions could be used to define more complex measures of repair strategies, and to determine the repair sequence and whether to ask questions. All of these works focus on globally and statically seeking the best action sequence for a problem setting where the actions and questions are not repeated.

### E. Summary

In conclusion, researchers have realized the benefits of DBNs and utility theory, and have begun to apply them to user modeling and related applications. Current research in these areas, however, is limited to passive inference, mostly affect-insensitive, and in a static domain. Efficiency in user-state inference is usually not considered and the utility of an action does not usually vary over time. In the affective-state assessment for user modeling and assistance, the information from sensory modalities is not sufficient and must be integrated with high-level models of the user and the environment.

Our system currently aims at two objectives of nonintrusive and active user-state inference, and dynamic and active sensor selection. For the first objective, our target is to design silent agents that use the most informative and nonintrusive evidences to infer user's affect, to provide the user with accurate and active assistance in a pervasive and ubiquitous computing environment. For the second objective, the selection of sensors in such a system should not be done once and then forgotten, rather, it needs to be continuously and dynamically reevaluated. We focus more on refining sensors/questions dynamically using a local optimal strategy.

### III. CONTEXT-PROFILE-STATE-OBSERVATION MODEL

BNs are probabilistic graphical models representing joint probabilities of a set of random variables and their conditional independence relations [19]. The nodes characterize the hypothesis/goal variables, hidden state variables, and evidence/observation variables of a physical system, while the arcs linking these nodes represent their causal relations. Hypothesis nodes represent what we want to infer while the evidence nodes represent sensory observations. The intermediate hidden nodes are necessary to model the state generation process. They link the hypothesis nodes with the observation nodes and therefore influence both the variables we observe and the variables we want to infer. Nodes are often arranged hierarchically at different levels, representing information at different levels of abstraction.

Static BNs (SBNs) work with evidences and beliefs from a single time instant. As a result, SBNs are not particularly suitable in modeling systems that evolve over time. DBNs have been developed to overcome this limitation. In general, a DBN is made up of interconnected time slices of SBNs, and the relationships between two neighboring time slices are modeled by an HMM, i.e., random variables at time  $t$  are affected by the variables at time  $t-1$ , as well as by the corresponding random variables at time  $t-1$  only. Fig. 1 illustrates such behaviors. The slice at the previous time is used in conjunction with current sensory data to infer the current hypothesis. DBNs represent a generalization of the conventional systems for modeling dynamic events, such as Kalman filtering and HMMs.

BNs have several advantages for modeling and inferring user's affective state. First, BNs provide a hierarchical framework to systematically represent information from different modalities at different levels of abstraction and systematically account for their uncertainties. Furthermore, with the dependencies coded in the graphical model, BNs can handle situations where some data entries are missing. Second, the user's dynamically changing state and the dynamic surrounding situations call for a framework that not only captures the beliefs of current events, but also predicts the evolution of future scenarios. DBNs provide a very powerful tool for addressing this problem by providing a coherent and unified hierarchical probabilistic framework for sensory information representation, integration, and inference over time. Furthermore, DBNs provide us with the ability to predict the influence of possible future actions through its temporal causality. Third, in many applications,

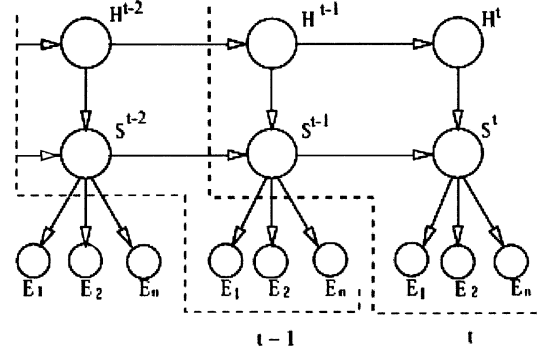


Fig. 1. Generic DBN consisting of three time slices, where  $H$  represents a collection of hypothesis nodes,  $S$  is a collection of hidden nodes,  $E_s$  is a collection of observation nodes, and  $t$  represents time.

the cost in terms of time, computational complexity, the interruption to the user, and the expense of information retrieval from various sensors, puts strict constraints on implementing the action decisions. DBNs provide facilities to actively and efficiently determine the utility of each sensory action and assistance over time.

Our generic framework to apply BNs to user modeling is the Context-Affective State-Profile-Observation model. It is used to infer users' affective state from their observations. As in Fig. 2, such model captures the user's profile, affective state, and the contextual information.

- *Context.* This component represents information about the specific environmental factors that can influence the user's affective state.
- *Affective state.* This component represents the user's emotional status. It constitutes the hypothesis we want to infer. Its values are different affective states, typically including fatigue, confused, frustration, fear, sad, and anger.
- *Profile.* This component model user's ability and competitiveness in finishing the operations. This provides the adaptation capability of the model to individual users.
- *Observation.* This component consists of sensory observations of different modalities characterizing user behaviors.

The affective state of the user and the hidden nodes of the user's visual, audio, and behavioral status in current time slice are influenced by the corresponding variables in the most recent time slice. The user profile could also have temporal links between time slices. However, in our model, we assume it remains unchanged. This figure also outlines the causal relations between context, profile, state, and observation variables as represented by arrows. The context and profile variables influence the user's affective states, while the user's affective states change the observation.

### IV. ACTIVE USER'S STATE INFERENCE

Purposive and sufficing information collection and integration are needed to infer about the user's affective state in a timely and economic manner. We are interested in how to dynamically control (select actions and make decisions) the system that has a repertoire of sensors such that the system operates in a purposive manner. We collect the observations from a subset of most

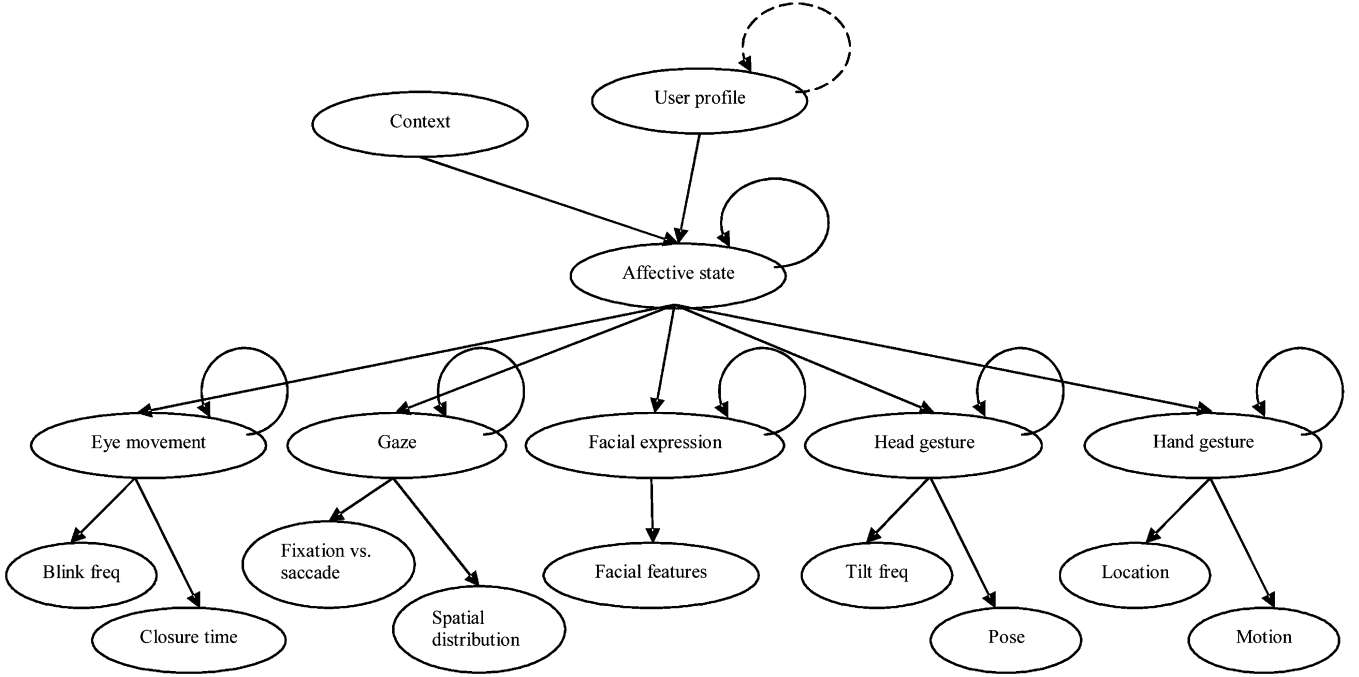


Fig. 2. Context-Profile-State-Observation model, where self-pointing arrows indicate temporal links.

informative sensors in order to recognize the user's affect efficiently and timely.

Mathematically, the user-affective state-inference problem may be viewed as a hypothesis-testing problem, with hypothesis,  $H = \{h_1, h_2, \dots, h_n\}$ , representing the possible user affective states. The sensory observation  $E$  is from  $m$  diverse sensors, i.e.,  $E = \{E_1, E_2, \dots, E_m\}$ . The goal is to estimate the posterior probability that  $H = h_i$  is true given  $E$ , i.e.,  $p(H = h_i | E)$ . According to the Shannon's measure of entropy, the entropy over the hypothesis variable  $H$  is:

$$\text{ENT}(H) = - \sum_i p(h_i) \log p(h_i). \quad (1)$$

$\text{ENT}(H)$  measures the uncertainty associated with  $H$ . It is zero when  $H$  is unambiguous, i.e., when one state has a probability of 1; it has the highest value when the probability is evenly distributed. The benefit of a piece of evidence to a hypothesis is measured by its expected potential to reduce the uncertainty with the hypothesis. This benefit can be quantified by mutual information, the differential entropy before and after the sensory action. Given the beliefs in hypothesis for the last time slice  $H^{t-1}$ , the mutual information of a sensory evidence  $E_j$  to current hypothesis  $H^t$  can be denoted as  $I(H^t; E_j)$

$$\begin{aligned} I(H^t; E_j) &= \text{ENT}(H^t) - \sum_j p(e_j) \text{ENT}(H^t | e_j) \\ &= - \sum_i p(h_i^t | H^{t-1}) \log p(h_i^t | H^{t-1}) \\ &\quad + \sum_j \left[ p(e_j) \sum_i p(h_i^t | H^{t-1}, e_j) \right. \\ &\quad \left. \times \log p(h_i^t | H^{t-1}, e_j) \right]. \end{aligned} \quad (2)$$

The above formula is fundamental for dynamically computing the uncertainty reducing potential for  $H$  due to  $E$ . We could easily extend it to consider the case that multiple sensors,  $E = \{E_1, \dots, E_n\} \subseteq E$ , are instantiated simultaneously

$$\begin{aligned} I(H^t; E_1, \dots, E_n) &= - \sum_i p(h_i^t | H^{t-1}) \log p(h_i^t | H^{t-1}) \\ &\quad + \sum_{e_1} \dots \sum_{e_n} \left[ p(e_1, \dots, e_n) \right. \\ &\quad \times \sum_i p(h_i^t | H^{t-1}, e_1, \dots, e_n) \\ &\quad \left. \times \log p(h_i^t | H^{t-1}, e_1, \dots, e_n) \right]. \end{aligned} \quad (3)$$

The probabilities in the above equation are readily available from the forward and backward inference propagation based on hypothesis beliefs for last time slice.  $p(h_i^t | H^{t-1}, e_1, \dots, e_n)$  is the posterior probability of hypothesis state for current time slice given a configuration on sensor states and the beliefs in hypothesis in the last time slice.  $p(h_i^t | H^{t-1})$  is the posterior probability of hypothesis state without acquiring new sensor evidence. In sensor selection,  $\text{ENT}(H^t)$  has the same value for all sensors and need not be calculated.

Acquiring information incurs cost. The cost may include the cost of information retrieval, the time to include the information from source into the fusion system, the computation time for sensory data processing, and the hardware execution time. We consider the sensor cost  $C$  of selecting  $E$ , a set of  $n$  sensors, where the costs for different sensors are assumed to be incorporated with the same importance, using the following formula:

$$C(E) = \sum_{i=1}^n C_i / \sum_{j=1}^m C_j \quad (4)$$

where  $C_i$  or  $C_j$  is the cost to acquire the information from sensor  $i$  or  $j$  and  $m$  is the total number of sensors. Combining the uncertainty-reducing potential and information-acquisition cost, we form the expected utility given sensor set  $E$  for current hypothesis  $H^t$  as:

$$EU(H^t, E) = \alpha I(H^t; E) - (1 - \alpha)C(E) \quad (5)$$

where  $\alpha$  is the balance coefficient between the two terms. The optimal sensor action can be found by using the following decision rule:

$$E^* = \arg \max_E EU(H^t, E). \quad (6)$$

We search for the best sensory action by examining the utilities for all configurations of sensors. Equation (6) is the fundamental equation for our dynamic and active sensing strategy. It allows our system to dynamically select a subset of sensors of the highest utility to current hypothesis in order to timely and efficiently estimate user's affect.

## V. DECISION ON ASSISTANCES

There are two key questions to answer considering the decision on assistance.

- 1) When should we provide assistance?
- 2) What assistance should we provide?

The first question normally requires a control threshold based on the probability distribution of the affective state variables. Thus, we calculate a state level (SL).

$$SL = \sum_i w(h_i^t) p(h_i^t) \quad (7)$$

where  $w(h^t)$  is the weight for an affective state, indicating this state's assistance level and  $p(h_t)$  is the posterior probability of the state after evidence propagation. More weight may be assigned to a more negative state such as fatigue. We can then set an engaging threshold (ET) on SL. If SL is greater than ET, we engage in assistance for users. In our experimentation, the SLs are smoothed over three time slices.

The type of assistance to provide depends on the utility of assistance. The utility of assistance represents the optimal tradeoff between its benefit and cost. The benefit focuses on beneficial consequence of the assistance. One measure of benefit is its potential to return the user from an anomaly to a normal state. It could be calculated by assessing the cross product of the situations and these assistances, through psychological experiments on a population of users or some assessment tools like unidimensional or multidimensional scaling. The cost includes the computational cost, the potential of annoying the user, the physical cost, and the cost of not providing or delaying the assistance.

The utility of assistance is also impacted by the user's current status, including the affective state, the current task goal, the cause, and the user's tolerance to assistance, shown in Fig. 3. "Task" shows the user's current interest, such as choosing some icon or button. "Cause" is the explanation for the subject's state. "Tolerance" is the intervention degree the user would agree on. In this paper, we only consider the impact of the affective state.

Let  $A_j$  represent the  $j$ th assistance in consideration. Let  $G_B(A_j, h_i)$  and  $G_C(A_j, h_i)$  represent the benefit and cost of

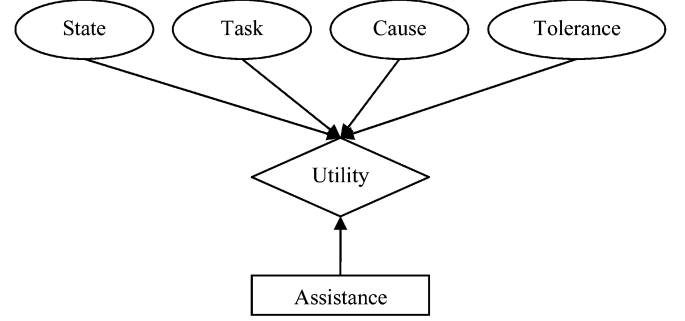


Fig. 3. Assistance utility model based on user's current status, task, cause, and tolerance of assistance.

assistance  $j$ , respectively, given user's current state  $h_i$ . Probabilistically, the benefit in form of the potential to return the user from an anomaly state  $h_i$  to the normal state  $h_0$  may be defined as  $G_B(A_j, h_i) = p(h_0 | A_j, h_i)$ . Then, the utility of assistance  $A_j$  given the current beliefs in hypotheses may be defined as

$$EU(A_j) = \sum_i G_B(A_j, h_i) p(h_i) - \sum_i G_C(A_j, h_i) p(h_i). \quad (8)$$

Similarly, benefit and cost should be scaled to maintain the same value range. The best assistance is determined via

$$EU^* = \arg \max_{A_j} EU(A_j). \quad (9)$$

Providing accurate and timely assistance remains a much more complex task in practice, largely due to the involvement of a human subject in the loop. The above strategy is a simple way to control assistance-engagement decisions. More complicated approaches, e.g., building up a tolerance-to-interruption model of the subject, or a closed-loop feedback control scheme, could be developed to achieve better effect.

## VI. EVALUATION

In this section, we present the evaluation results of our framework using both simulated data and real data. For simulated study, the task in evaluation is to detect whether a computer operator is among three affective states: fatigue, nervousness, and confusion, using various visual cues. For the experiment with real data, we applied our framework to detect and monitor human fatigue.

### A. Experiments With Synthetic Data

First, we evaluate the feasibility of our framework and the effectiveness of the proposed active sensing strategy using synthetic and subjective data. The model implementation is in MATLAB using the BNT toolkit [18]. The inference algorithm we use is the junction tree engine.

1) *Experimental Model and Parameters:* Fig. 4 shows the DBN model we used. Table I summarizes the discrete variables and their states used in the evaluation model. We use three separate hypothesis nodes for three affective states (fatigue, nervousness, and confusion) because we do not require that these states are exclusive from each other. The assistance node here is to show the impact of the chosen assistance on affective states, although very hard to estimate in practice.

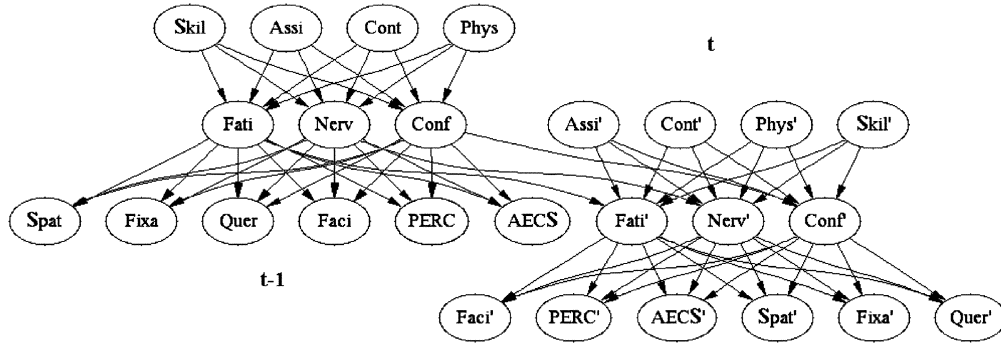


Fig. 4. DBN network structure used in evaluation, which uses five visual cues and direct query to assess three affective user states.

TABLE I  
VARIABLES AND THEIR STATES USED IN THE EVALUATION MODEL

Component	Variables	States	Description
Assistance	Assistance	null/warning/emphasis/simplification	
Context	Context	complex/simple	The surrounding environment of the subject.
Profile	Physical condition	strong/weak	Physical condition of the subject.
	Skill	strong/weak	Computer skill of the subject.
Affective states	Fatigue	positive/negative	
	Nervous	positive/negative	
	Confusion	positive/negative	
Sensor	Facial expression	neutral, happiness, sadness, anger, surprise, disgust, fear	Expressions from the FACS system. (1)
	Eyelid PERCLOS	high/normal/low	Percentage of eyelid closure over time. (2)
	Eyelid AECS	fast/normal/slow	Average eye closure/open speed over time. (3)
	Gaze spatial distribution	high/even/low	The ratio of gaze fixations on computer to gaze fixations outside computer over time. (4)
	Gaze fixation ratio	high/even/low	The ratio of gaze fixation to saccade eye movement over time. (5)
	Query	fatigued/confused/no	Answers to direct questioning in the form "are you ...?" among "fatigued, confused, comfortable." (6)

The parameters of our BN model include the prior probabilities for context and profile nodes, and the conditional probabilities for the links. Since this study focuses more on the working mechanism than a fidelity model, most of the required probabilities were specified manually. In particular, the prior probabilities are all three affective state nodes set to (0.5, 0.5). The transitional probabilities between affective states in two consecutive time slices are specified accordingly. For example, the transitional probability between the same states of two slices, e.g., positive to positive for fatigue, is high, if we consider a user's mental state remain relatively stable. For transient affect such as confusion, which may come and go quickly, the transitional probability may be lower. The transitional probability between opposite states, positive to negative or negative to positive, is much lower correspondingly. In our experimentation, the transitional probability between the same states of fatigue is 0.9, while it is 0.85 for the two other affective states. Other condi-

TABLE II  
SIX DIFFERENT SETTINGS USED IN EVALUATION

Setting	Notes
Passive Fusion I	Randomly select 1 sensor and select the same sensor.
Active Fusion I	Actively select 1 sensor.
Passive Fusion II	Randomly select two sensors at each time slice.
Active Fusion II	Actively select two sensors at each time slice
Assistance	Provide assistance when in active fusion II.
Various Costs	Use different sensor costs in calculating the utility values of sensors.

TABLE III  
BENEFIT AND COST OF ASSISTANCE TO AFFECTIVE STATES

Assistance	Benefit ( $G_B$ )			Cost ( $G_C$ )		
	Fatigued	Nervous	Confused	Fatigued	Nervous	Confused
Warning	0.56	0.33	0.13	0.07	0.33	0.57
Emphasis	0.25	0.60	0.31	0.43	0.07	0.36
Simplify	0.19	0.07	0.56	0.50	0.6	0.07

tional probability values are obtained subjectively in a similar fashion. For those probabilities that are hard to estimate manually, we apply the noise-or principle or its extension [19]. The noise-or principle allows estimating the joint conditional probabilities based on the marginal conditional probabilities, therefore, greatly reducing the number of probabilities to estimate.

2) *Experimental Scenarios, Settings, and Data:* During the experiment, we compare the results of different sensor activation and assistance strategies, covering aspects of passive and active state inference with/without assistance process. These settings are listed in Table II. Among the first five settings, the sensor costs are all set as zero in active fusion, i.e.,  $\alpha$  is 1 in (5). In the last setting, we use a set of nonzero sensor costs.

In assistance setting, all ETs are set to 0.8. The benefits and costs of assistance are subjectively assigned. Table III summarizes the types of assistances, their benefits and costs with respect to each user affective state. Query provides a more accurate estimate of the user's affective state. However, it is intrusive to the subject and thus is associated with a high cost. We demonstrate the function of query by only using it as the last confirmation before any assistance. When the individual or overall SLs exceed the predefined ET, an answer from the query channel is retrieved to update the beliefs in the subject. If after this, the SL is still significant, the utilities for assistances are calculated and the one with the highest value is chosen to instantiate the assis-

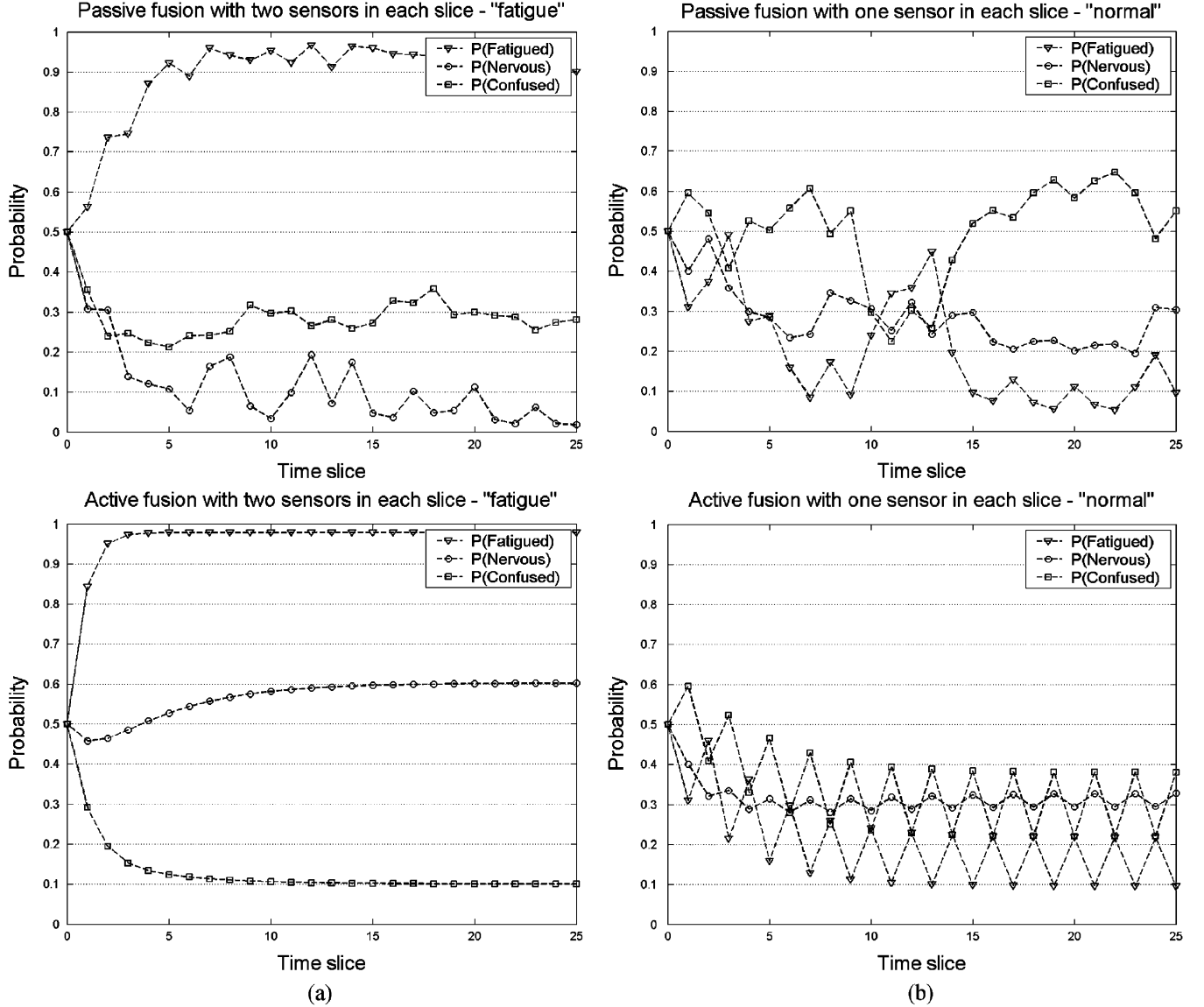


Fig. 5. Passive fusion with random selection versus active fusion, with one and two sensors activated. (a) Fatigue. (b) Normal.

tance node. We assume the assistance is engaged in for certain time duration (five slices in our experimentation). We simulate and observe the effect by the causal link between the affect state nodes and the assistance node.

Data for different affective state scenarios (fatigue, nervousness, confusion, and normal) were synthetically generated. In generating data for each scenario, we first instantiate the corresponding affective hypothesis node to its desired state with certain probability and then perform a forward propagation to determine the probability distribution for each sensor. For example, for the scenario where the subject is fatigued, the fatigue node is instantiated with a 99% probability for the positive state and 1% for its negative state; also the probabilities for the positive states of nervousness and confusion are set to 1% and negative states 99%, respectively. After forward propagation of beliefs from the affective state nodes to the sensor nodes, each sensor has a probability distribution associated with its states. Then in the generated data of each scenario, each sensor turns out the state of the highest probability among all possible states,

the state most indicative of the underlying affective status. In other words, such sensory channel could always catch the most likely expression resulted from the underlying affective status.

3) *Experimental Results:* We use the six settings for the three affective scenarios, i.e., fatigue, nervousness and confusion, and the normal scenario, fed with the state generated above to instantiate the selected sensors in each time slice. The posterior probability for the positive state of each affect variable in each time slice is recorded, as well as the calculated SLs. Thereafter, we call this probability the belief of the corresponding affect, e.g., the belief of fatigue. This belief and the information entropy associated with the probability distribution are the measure in evaluating various settings. We summarize the results of comparison of active fusion versus passive fusion without/with assistance.

Fig. 5 shows the belief curves for fatigue and normal scenarios with one or two sensors are selected in each time slice respectively. The sensors are selected randomly in the passive fusion. As the curves show, active fusion (on the right) detects

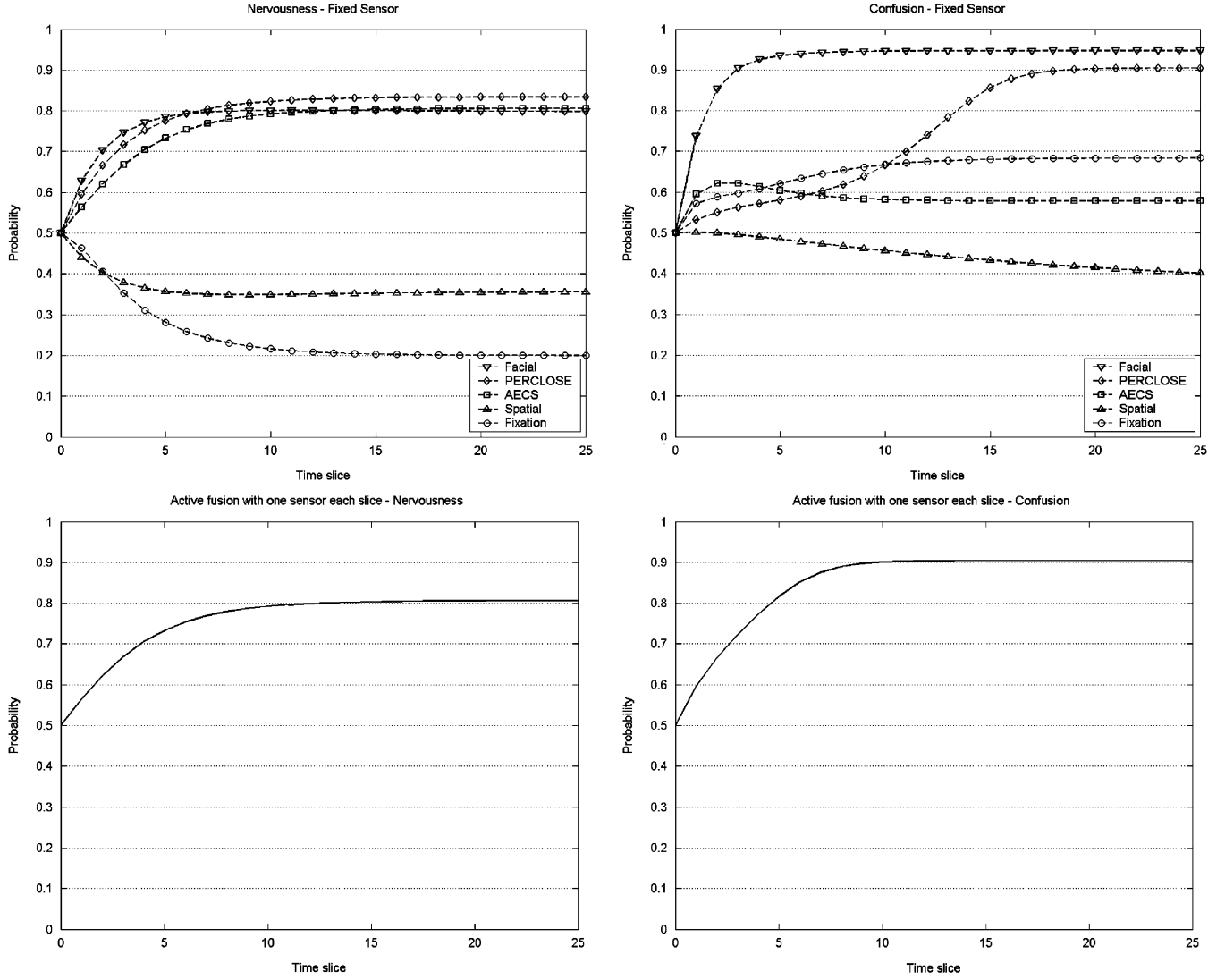


Fig. 6. Active fusion versus passive fusion with fixed sensor selection, with one sensor activated.

the underlying status of the subject more quickly. Specifically, in fatigue scenario, when two sensors are selected in each time slice, the belief of fatigue rises fast to around 1 while the beliefs of other affects remain relatively low. In fact, the results for nervousness and confusion scenarios show the same features. In normal scenario, all three probabilities drop below 0.5. Although the corresponding passive fusion settings could detect the same trends in these beliefs, they are not as efficient as active fusion.

Now, we compare active fusion versus the passive fusion, where a fixed sensor is selected. Fig. 6 shows the performance of selecting each of the visual sensors 1–5, respectively, for the nervousness and confusion scenarios. For each scenario, we only show the belief change for the positive state of the underlying selective state, i.e., nervousness and confusion in the two scenarios, respectively. From the curves in the passive fusion, such belief could yield a very bad value when using certain sensors. For example, using spatial or fixation in the nervousness and using spatial in the confusion scenario even produce a belief below 0.5 for the underlying affect. However, the active fusion can consistently outperform the passive inference, yielding the

TABLE IV  
NUMBER OF SLICES NEEDED TO REACH THE TARGET THRESHOLD ON SL FOR AFFECTIVE SCENARIOS, WHERE n/a INDICATES THE THRESHOLD IS NEVER REACHED DURING THE 25 TIME SLICES

	One sensor (threshold = 0.8)		Two sensors (threshold = 0.9)	
	Passive	Active	Passive	Active
"Fatigue"	7	5	7	5
"Nervousness"	n/a	14	n/a	14
"Confusion"	16	6	16	6

inference performance comparable to the best in the fixed sensor selection. Similarly, the belief changes for other affective states further confirm the above observation.

We could also examine the uncertainty reduction abilities by setting a target threshold on the underlying affect belief, and compare in different settings the number of time slices needed to first reach this threshold. This is a good measure, since in practice we could regard this threshold as the control threshold for assistance engagement. Here we set the target threshold as 0.8 for activating one sensor and 0.9 for two sensors. The results are summarized in Table IV. It is apparent that for both one sensor



TABLE V  
SENSOR SEQUENCES IN EACH SCENARIO IN ACTIVE FUSION WHERE ONE  
SENSOR IS SELECTED IN EACH TIME SLICE, WITH (a) NO SENSOR COST AND  
(b) LARGER COST FOR SENSORS WITH HIGHER MUTUAL INFORMATION VALUE

Time slice	Fatigued (a)	Fatigued (b)	Nervous (a)	Nervous (b)	Confuse d (a)	Confuse d (b)	Normal (a)	Normal (b)
1	3	3	3	3	3	3	3	3
2	3	3	3	3	2	2	2	2
3	3	5	3	3	2	2	5	5
...	...	...	...	...	...	...	...	...
13	3	5	3	3	2	2	5	5
14	3	5	3	2	2	2	2	2
...	...	...	...	...	...	...	...	...
25	3	5	3	2	2	2	5	5

and two sensor cases, active fusion usually takes less time to reach assistance ET. This demonstrates that active sensing allows providing assistance in a timelier manner than the passive sensing.

In the paragraphs below, we want to examine the sensor sequences selected in active fusion with and without costs. Active fusion selects the sensors with the highest utility in each time slice. This utility may change over time, even in the same scenario. Table V shows the sensor sequences for different scenarios with or without sensor costs. Because the way we assign the initial beliefs, the first sensor selected for all scenarios are all the same, AECS (sensor 3). Subsequently, with the change of affective hypothesis beliefs, different sensors may be selected, based on their utility. However, we notice that not all sensors are used. Specifically, only sensors 2 (PERCLOS), 3 (AECS), and 5 (gaze fixation ratio) are ever selected in all scenarios. We also notice that in the later time slices, the sensor sequence is fixed, with certain sensors repeatedly selected.

We further investigate the impact of sensor cost on inference performance. In Fig. 7, the curves of beliefs for those affect variables show difference for the two active fusion settings in the fatigue scenario. As shown in Table V, different sensors are selected due to the impact of sensor costs. In the setting on the right that assigns sensor costs, sensor 5, gaze fixation ratio, since it has a lower cost, is selected over sensor 3, AECS. Although the mutual information value of gaze fixation ratio is not the highest, the evidence from this sensor yields better beliefs for the underlying affect hypotheses in this scenario. The belief of fatigue is higher while the belief of nervousness is much lower. This reminds us that the mutual information is just an expected benefit calculated on the probability distribution of hypotheses. A sensor with the highest value of mutual information is not absolutely superior to others. As a result, we should not be too rigid in using this mutual information.

Finally, we want to evaluate the effectiveness of the proposed assistance scheme. Although the corresponding individual SL for the underlying affect in each “affective” scenario reaches the threshold (0.8 here) very fast, the assistance is available after the fifth time slice since we want to be cautious to accumulate evidence for enough time. For each scenario, we calculate the utility for each assistance type and choose the assistance with the highest utility. The SLs and corresponding utilities for dif-

ferent assistances when the assistance is engaged in are given in Table VI.

The probabilities for affective states in assistance process are shown in Fig. 8. In our case, the assistance is appropriate for all scenarios, i.e., warning for fatigue scenario, emphasis for nervousness scenario for the subject to focus, and interface simplification for confusion scenario. Then, the chosen assistance is used to instantiate the corresponding assistance state in the model for a certain number of time slices. While the assistance soothes the danger of one affect, it may aggravate other affects, e.g., warning may intensify confusion. This tells us the importance of accuracy in detecting the subject’s status in providing appropriate assistance.

### B. Experiments With Real Data

In this section, we present results of evaluating the effectiveness of the proposed affect recognition framework for real-time human fatigue detection. To validate our fatigue model, we perform a human subject study. The study includes a total of eight subjects. Two test bouts are performed for each subject. The first test is done when they first arrive in the lab at 9 PM and when they are fully alert. The second test is performed early in morning about 7 AM the following day, after the subjects have been deprived of sleep for a total of 25 h. During the study, the subjects are asked to perform a test of variables of attention (TOVA). The TOVA consists of a 20-min psychomotor test, which requires the subject to sustain attention and respond to a randomly appearing light on a computer screen by pressing a button. The response time is used as a metric to quantify the subject’s performance and, in comparison, as the ground truth baseline of the human subject’s fatigue level.

A more complicated BN model has the fatigue as the hypothesis, contextual and profile nodes such as the human subject’s physical condition and time circadian, and a set of similar visual sensory nodes about eyelid movement, head movement, and facial expression as given in Table I. The model parameters are obtained from a combination of different sources: subjective data, surveys, and training data. A computer vision system uses a remotely located CCD camera to acquire video images of the human subject. Visual cues are extracted in real time and combined to infer a composite fatigue score of the driver. We study the validity of this composite fatigue score (probability of fatigue) estimated from the model versus the TOVA performance.

Fig. 9 plots the estimated fatigue score and the TOVA performance (as measured by response time) over time. It clearly shows that the fatigue score curve correlates well with response time curve, therefore proving the validity of the fatigue score in quantifying fatigue and performance. More details on this study about the vision techniques to extract those visual cues, the BN model structure and parameters, and the experimentation setting and result are available in [15].

### C. Discussion

In our evaluation experiments, we use a dynamic Bayesian model to assess several typical human mental affects and use a set of data to compare different fusion strategies. We could draw several conclusions from the results.

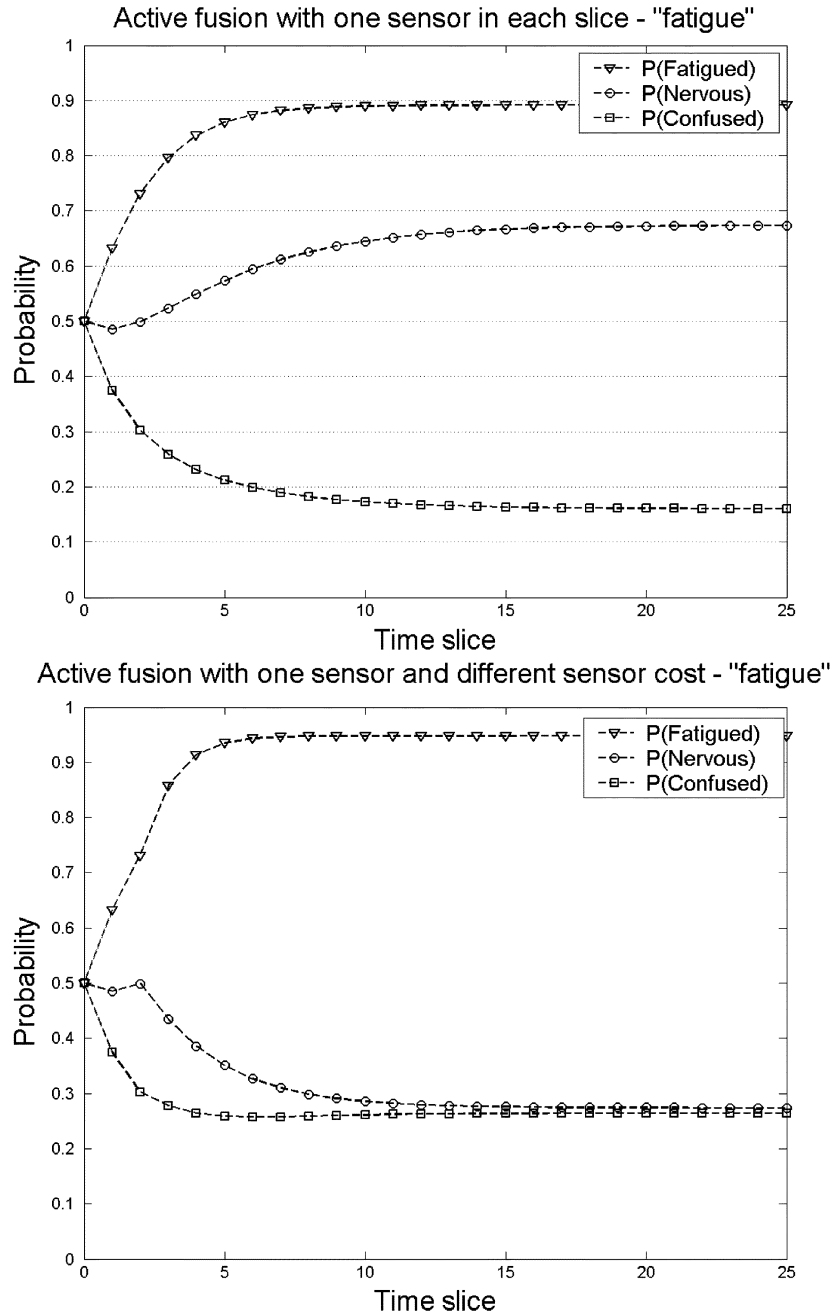


Fig. 7. Belief curves for fatigue scenario, where different sensor costs change the sensor selection sequence in active fusion.

TABLE VI  
SLS AND ASSISTANCE UTILITIES IN AFFECTIVE SCENARIOS,  
WITH  $ET = 0.8$  AND ACTIVATING TWO SENSORS

Scenario	SL			Utility		
	Fatigued	Nervous	Confused	Warning	Emphasis	Simplification
Fatigued	0.87	0.53	0.19	0.34	0.13	-0.46
Nervous	0.10	0.85	0.75	-0.29	0.40	-0.12
Confused	0.07	0.12	0.81	-0.33	0.02	0.31

1) In most cases, especially at the beginning stage, active fusion is efficient in building up a pretty good belief distribution for the underlying status of the subject. The update of affect beliefs distinguishes the different affect states very

fast. Such ability in differentiating various states is very useful and important for practical systems.

- 2) Normally, selecting more sensors simultaneously, as shown in the result using two sensors at each time slice, could provide more accuracy and better performance. This reasonably shows the advantage of multisensor fusion. On the other hand, we have to consider the cost of using more sensors, and even the cases where sensors could not be activated together. This is trade-off issue, not just the more the better.
- 3) Different sensor costs change the selected sensor sequence. This indicates that in practice, the constraint on sensor's cost may impact the performance. We also notice that mutual information is not an absolute mea-

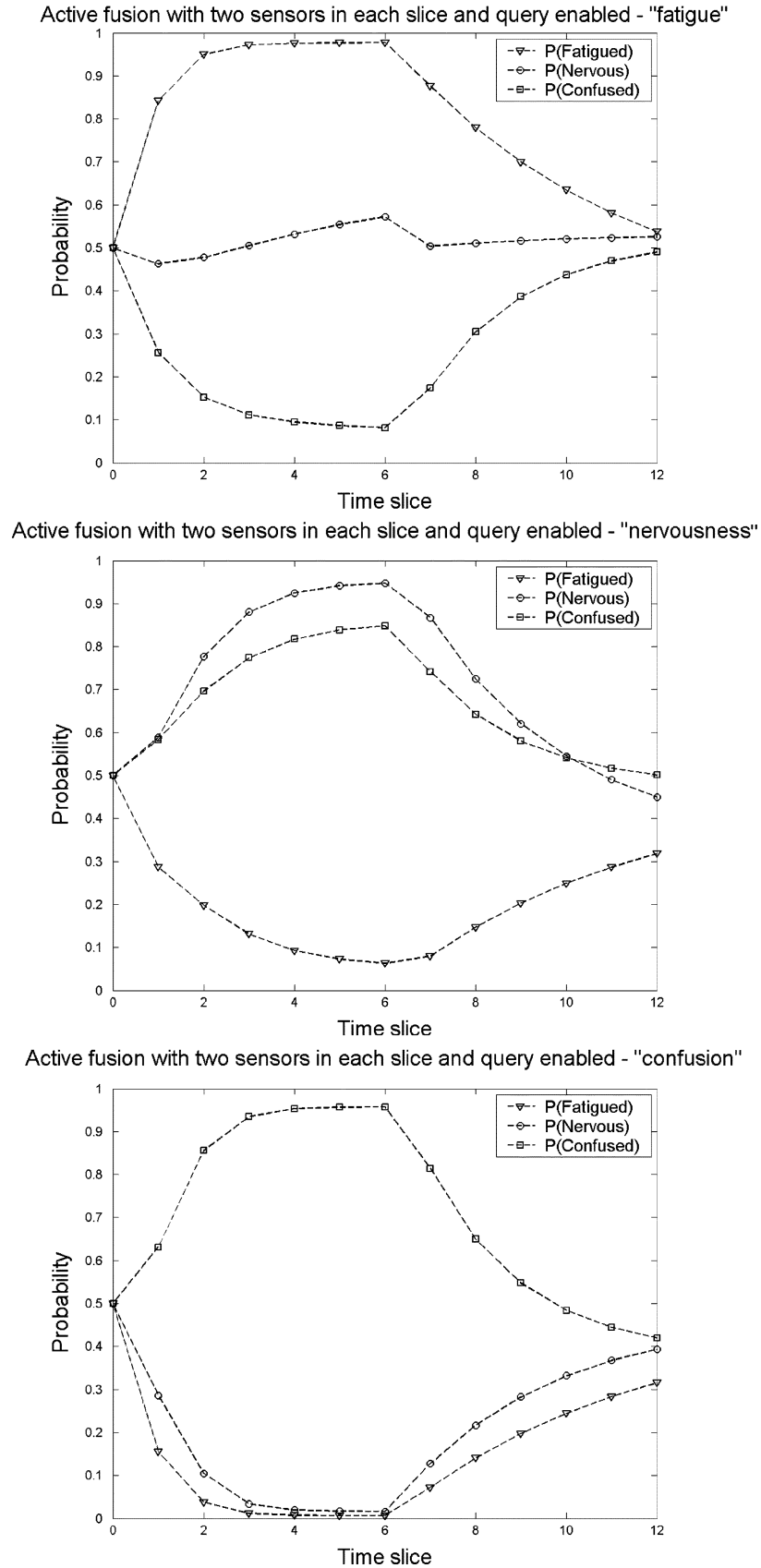


Fig. 8. Assistance process shown for the three "affective" scenarios, with two sensors activated in each time slice, and the ET set at 0.8.

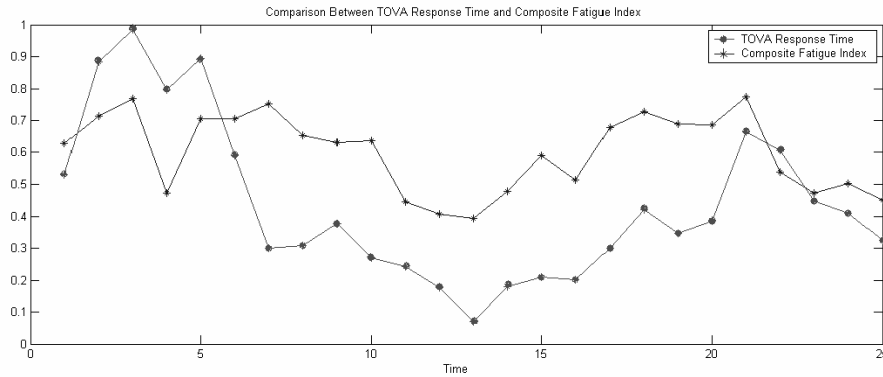


Fig. 9. Estimated composite fatigue score versus the normalized TOVA response time. The two curves track each other well. It is clear that the two curves' fluctuations match well, proving their correlation and covariation.

sure, as shown in some results. Selection strategies that incorporate randomness in selection process may even improve performance.

- 4) The experimentation also reveals a side effect of active sensing. After the initial buildup stage, inference degradation might occur, where some sensors dominate the active sensor selection and get selected repeatedly.

## VII. CONCLUSION

Our research aims at dynamically and systematically modeling the user affective state, and performing active information fusion so that the user's state and need can be determined and met in a timely and efficient manner. In our study, a generic framework based on DBN is built to account for various modalities in user modeling; an active information fusion strategy using information theory is proposed to assess user's affective states; and decisions on appropriate assistance could be evaluated based on utility theory incorporating affect beliefs. In particular, the framework provides a mechanism for dynamically selecting the best subset of sensors, according to a tradeoff between sensor costs and the expected information they can provide. A set of experiments involving both synthetic and real data demonstrate the feasibility of the proposed framework as well as the effectiveness of the proposed active sensing strategy for quick and efficient decision making.

Our contribution can be summarized in three areas: 1) systematically model the uncertainty, dynamics, and different types of knowledge associated with user affective state using DBN; 2) propose information-theoretic mechanism to perform active and purposive user affective state inference in a timely and efficient manner; and 3) propose information-theoretic criteria to optimally determine when and what assistance to provide to maximize the chance for returning the user to its normal state while minimizing interference with the user's work and user's annoyance. The main contribution lies in the integration of the proposed active sensing mechanism into the DBN-based framework for user state inference and user assistance.

This work advances the theory and application of efficient information fusion in human-computer interaction, especially in the cutting edge research areas of pervasive and ubiquitous

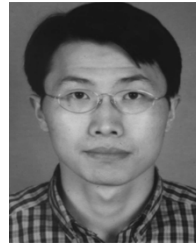
computing. It will improve user's performance and productivity by augmenting human cognition negatively affected by adverse affects. Furthermore, it will also identify deficiencies in the interface, minimize accidents caused by human errors; and improve human machine interaction experience. A broad range of applications will exist for the proposed research including military for improving warfighters' interaction with computer based systems, especially under stressful operational environments; transportation for enhancing drivers and pilots' performance to improve safety; improved decision aiding systems for individual and team settings in critical, typically high-stress applications such as air traffic control, process control in nuclear power plants and chemical plants, and emergency vehicle dispatchers; assistance to people with disability.

We notice that such an affective state-detection system alone could not fully fulfill very accurate assistance. We make such a statement since we observe that even with the carefully designed working procedures and paradigms, such a single assessment model could not in some cases recognize the status of the subject very accurately and thus might fail to provide urgent assistance. This is especially true when we consider the variability of individual personality, the configuration complexity for the large number of node states, and especially the strict requirement of accuracy on such assistance systems. Further research is ongoing in our lab to integrate multiple and heterogeneous models in such task to improve the robustness and performance of user state detection and assistance decision.

## REFERENCES

- [1] T. Bohnenberger, B. Brandherm, B. Großmann-Hutter, D. Heckmann, and F. Wittig, "Empirically grounded decision-theoretic adaptation to situation-dependent resource limitations," *Künstliche Intell.*, vol. 3, pp. 10–16, 2002.
- [2] G. Ball and J. Breese, "Emotion and personality in a conversational agent," in *Embodied Conversational Agents*. Cambridge, MA: MIT Press, 2000.
- [3] M. Bauer, P. Gymtrasiewicz, and J. Vassileva, *User Modeling*. New York: Springer-Verlag, 2001.
- [4] I. Cohen, A. Garg, and T. S. Huang, "Emotion recognition using multi-level-HMM," in *NIPS Workshop Affective Comput.*, Denver, CO, 2000.
- [5] C. Conati, "Probabilistic assessment of user's emotions in educational games," *Appl. Artif. Intell.*, vol. 16, no. 7–8, 2002.
- [6] DARPA Augmented Cognition Program (2004, May). [Online]. Available: <http://www.augmentedcognition.org/>

- [7] M. W. Hass and L. Hettinger, *Adaptive User Interfaces*. Mahwah, NJ: Lawrence Erlbaum, 2001.
- [8] D. Heckerman, J. S. Breese, and K. Rommelse, "Decision-theoretic: Troubleshooting," in *Commun. ACM*, vol. 38, 1996, pp. 49–57.
- [9] E. Horvitz, "Uncertainty, action, and interaction: In pursuit of mixed-initiative computing," *IEEE Intell. Syst.*, no. 6, pp. 17–20, Sep./Oct. 1999.
- [10] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse, "The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users," in *Proc. 14th Conf. Uncertainty Artif. Intell.* San Francisco, CA, 1998, pp. 256–265.
- [11] E. Horvitz and T. Paek, "A computational architecture for conversation," in *Proc. 7th Int. Conf. User Modeling*, New York, 1999, pp. 201–210.
- [12] —, "DeepListener: Harnessing expected utility to guide clarification dialog in spoken language systems," presented at the 6th Int. Conf. Spoken Lang. Process., Beijing, China, 2000.
- [13] E. Hudlicka and M. D. McNeese, "Assessment of user affective and belief states for interface adaptation: Application to an air force pilot task," *User Modeling User-Adapted Interaction*, vol. 12, pp. 1–47, 2002.
- [14] A. Jameson, "Numerical uncertainty management in user and student modeling: An overview of systems and issues," *User Modeling User-Adapted Interaction*, vol. 5, no. 3–4, pp. 193–251, 1996.
- [15] Q. Ji, Z. Zhu, and P. Lan, "Real time nonintrusive monitoring and prediction of driver fatigue," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1052–1068, Jul. 2004, to be published.
- [16] H. Langseth and F. V. Jensen, "Decision theoretic troubleshooting of coherent systems," *Reliability Eng. Syst. Safety*, vol. 80, no. 1, pp. 49–61, 2003.
- [17] P. Maes and B. Schneidman, "Direct manipulation vs. interface agents: A debate," *Interactions*, vol. IV, no. 6, 1997.
- [18] K. Murphy, "The Bayes net toolbox for Matlab," *Comput. Sci. Stat.*, vol. 33, pp. 331–350, 2001.
- [19] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.
- [20] R. Picard, *Affective Computing*. Cambridge, MA: MIT Press, 1997.
- [21] R. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.
- [22] D. V. Pynadath and M. P. Wellman, "Accounting for context in plan recognition, with application to traffic monitoring," in *Proc. 11th Conf. Uncertainty Artif. Intell.*, 1995, pp. 472–481.
- [23] Y. Qi and R. Picard, "Context-sensitive Bayesian classifiers and application to mouse pressure pattern classification," presented at the Int. Conf. Pattern Recognition, Quebec City, Canada, 2002.
- [24] G. I. Webb, M. J. Pazzani, and D. Billsus, "Machine learning for user modeling," *User Modeling User-Adapted Interaction*, vol. 11, pp. 19–29, 2001.



**Xiangyang Li** (M'00) received the B.S. degree in automatic control from Northeastern University, Shenyang, China, in 1993, the M.S. degree in systems simulation from the Chinese Academy of Aerospace Administration, Beijing, in 1996, and the Ph.D. degree in information security from Arizona State University, Tempe, in 2001.

He was a Researcher with the Beijing Institute of Computer Application and Simulation Technology, Beijing, China, from 1996 to 1999. He is currently an Assistant Professor in the Department of Industrial and Manufacturing Systems Engineering, University of Michigan, Dearborn. He has published more than 20 papers in peer-reviewed journals and conferences. He participated in research projects funded by federal agencies including DARPA, AFOSR, ARO, and NIST. He has research interests in information system security/assurance and quality of service, user modeling and assistance in human systems studies, data mining in knowledge discovery and management, and system modeling and simulation.

Prof. Li is a Member of the Association for Computing Machinery, the Association for Information Systems, and the Chinese Association for Systems Simulation.



**Qiang Ji** (S'92–M'98–SM'04) received the M.S. degree in electrical engineering from the University of Arizona, Tucson, in 1993 and the Ph.D. degree in electrical engineering from the University of Washington, Seattle, in 1998.

He is currently an Associate Professor in the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. His areas of research include computer vision, probabilistic reasoning for decision making and information fusion, pattern recognition, and robotics.

Between May, 1993 and May 1995, he was a Research Engineer with Western Research Company, Tucson, AZ. In the summer of 1995, he was a Visiting Technical Staff Member with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, where he developed computer vision algorithms for robotics applications. In the summer of 2003, he was a Visiting Faculty Fellow with Information Directorate, Air Force Research Laboratory, conducting research on decision making under uncertainty. He has published more than 60 papers in peer-reviewed journals and conferences. His research has been funded by local and federal government agencies including the NSF, NIH, AFOSR, ONR, DARPA, and ARO, and by private companies including Boeing and Honda. His latest research focus is on applying computer vision and probabilistic reasoning theories to human computer interaction, including human-fatigue monitoring, user-affect modeling and recognition, and active-user assistance.