2013

Human desire inference process and analysis

Jeyoun Dong
Iowa State University

Human desire inference process and analysis

by

Jeyoun Dong

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Computer Science

Program of Study Committee:
Carl K. Chang, Major Professor
Johnny S. Wong
Ying Cai
Simanta Mitra
Debra Satterfield

Iowa State University

Ames, Iowa

2013

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis. First and foremost, Dr. Carl K. Chang, as my advisor, for his academic advice and guidance, patience and support throughout the whole research works and the writing of this thesis. His insights and words of encouragement have often raised me up, so I renewed my hopes and tided over big adversity for completing my graduate education.

I would like to deeply appreciate to Dr. Johnny Wong, Dr. Ying Cai, Dr. Simanta Mitra, and Prof. Debra Satterfield for their efforts to be my POS (program of study) committee members.

I would like to give a special thanks to all former and current group members in software engineering lab, CS Korean alumni, and CS department staffs for their helpful collaboration and discussion during my PhD, and all of my friends, more obviously, Lenwood Monte, Hyejin Ryu, Yeunsook Lee, On-Soon Lee, Hyuck Hur, Sungjun Kim, Hajeong Lim's family, and Dr. Minha Lee's family for their warm and valuable comments.

I would especially like to thank Hongwon Kang and members for their support and prayer over the years.

Finally, I am forever indebted to my dad, mom, brother's family (Changsun, Soyoung, and Hyunhye), and Myunghwan for instilling in me confidence and a drive for pursuing my PhD.

ABSTRACT

Ubiquitous computing becomes a more fascinating research area since it may offer us an unobtrusive way to help users in their environments that integrate surrounding objects and activities. To date, there have been numerous studies focusing on how user's activity can be identified and predicted, without considering motivation driving an action. However, understanding the underlying motivation is a key to activity analysis. On the other hand, user's desires often generate motivations to engage activities in order to fulfill such desires. Thus, we must study user's desires in order to provide proper services to make the life of users more comfortable.

In this study, we present how to design and implement a computational model for inference of user's desire. First, we devised a hierarchical desire inference process based on the Bayesian Belief Networks (BBNs), that considers the affective states, behavior contexts and environmental contexts of a user at given points in time to infer the user's desire. The inferred desire of the highest probability from the BBNs is then used in the subsequent decision making.

Second, we extended a probabilistic framework based on the Dynamic Bayesian Belief Networks (DBBNs) which model the observation sequences and information theory. A generic hierarchical probabilistic framework for desire inference is introduced to model the context information and the visual sensory observations. Also, this framework dynamically evolves to account for temporal change in context information along with the change in user's desire.

Third, we described what possible factors are relevant to determine user's desire. To achieve this, a full-scale experiment has been conducted. Raw data from sensors were interpreted as context information. We observed the user's activities and get user's emotions as a part of input parameters. Throughout the experiment, a complete analysis was conducted whereas 30

factors were considered and most relevant factors were selectively chosen using correlation coefficient and delta value. Our results show that 11 factors (3 emotions, 7 behaviors and 1 location factor) are relevant to inferring user's desire.

Finally, we have established an evaluation environment within the Smart Home Lab to validate our approach. In order to train and verify the desire inference model, multiple stimuli are provided to induce user's desires and pilot data are collected during the experiments. For evaluation, we used the recall and precision methodology, which are basic measures. As a result, average precision was calculated to be 85% for human desire inference and 81% for Think-Aloud.

CHAPTER 1. INTRODUCTION

The emergence of mobile and pervasive computing has fundamentally changed the interaction patterns between human users and computer systems. This trend places increased demands on the systems capabilities to satisfy each individual user. With untethered access anytime and anywhere, users expect computing systems to be smarter and more personalized, giving the requirement that the systems must behave adaptively. Based on contexts, systems anticipate ever-changing needs of users [1]. For instance, a computer system based smart home should gradually increase the frequency and level of detail of key reminders provided as the resident grows older and experiences degradation of memory functions.

To date, studies on human-computer interaction (HCI) have been mostly focusing on investigating user modeling and intelligent assistance system to understand, explain, and augment user actions [8, 9, 10]. Many existing context-aware applications such as user modeling, that traditionally focused on what is generally condensed "rational aspects of user behavior", are to design and implement autonomous agent to assist and make users feel more comfortable in their daily work and life. Such a model typically fails to adapt to user's basic affective states [2, 3, 18, 19, 20]. Alternatively, intelligent assistance systems provide the user with timely and appropriate assistance that captures interprets and responds to the internal states of the human user. Although intelligent assistance systems can be systematically planned to overcome uncertain and single sensory observation and user's changing mental states for the assistance, they do not go deeper to explore human desire that is actually essential to understanding human mental states [5, 30, 31].

However, human desire has long been identified as a philosophical problem, not a main subject in the HCI field in the past decades, even though it represents the motivational aspects of human behaviors [4, 42]. Desire (i.e., a sense of longing for something or someone) becomes stronger as the thought continues and eventually leads to objective actions to achieve some sort of goal.

The more the system knows what the user's desire is exactly, the more the human being can achieve an ultimate goal that they want. It is essential to improving the accuracy for providing appropriate system services to the user. Moreover, affective states should be included as one of important factors for inferring desire because an incorrect desire can be inferred without consideration of affective states [5, 42].

For example, assume that students are taking a lecture in a classroom. One student raises a hand in the classroom and the action can be interpreted in terms of the desire for "questions", "answers", "other purposes (e.g., ask to go restroom)", and so on. Student's desire would be individually different even if the student's behaviors are identical. There are examples including same activities and different desires. If the student raises a hand with happiness (satisfaction), her behavior can be interpreted as "understanding" or "knowing an answer to teacher's question". Thus, the desire of the student is to "answer the teacher's question." On the other hand, if she raises a hand with sadness (dissatisfaction), her behavior can be explained as "misunderstanding the lecture", and her desire can be interpreted as "asking a question about the lectures".

The main purpose of this study is to infer human desire by detecting internal mental states and what users desire and how to fulfill it. A probabilistic model based on the Bayesian Belief Networks (BBNs) is employed to formalize the inference process. First, a hierarchically organized probabilistic model for the desire inference is introduced to use the context information such as behaviors, environments, and affective states with the visual sensory observations. Second, we precisely define the belief-perceived situation as $B(m, a, e)$, and present how to use belief-perceived situations to infer user's desire. Emotions play a key role in human creativity [5, 30] and intelligence, as well as in human logical thinking and decision-making. In addition, emotions are used interchangeably with affection in some literature [57, 58, 59]. The proposed model includes seven affective states (i.e., happy, surprised, angry, disgust, fear, sad and neutral) based on the OCC (Ortony, Clore, & Collins) model [5]. Third, we extend a probabilistic framework based on the Dynamic Bayesian Belief Networks (DBBNs) that model the observation sequences (time series) assisted by information theory. A generic hierarchical probabilistic framework for desire inference is introduced to model the context information and the visual sensory observations. Also, this framework dynamically evolves to account for temporal changes in context information along with the change in user's desire.

Fourth, we conduct a pilot study to collect dataset for desire inference, to establish the correlation between factors and desire, and to identify the relevant factors for desire inference. Finally, we evaluated the desire inference computational framework using recall and precision methodology. They are basic measures used in evaluation. Thus, we have established an evaluation environment within the Smart Home Lab (SHL) to validate our approach. During the pilot study, data are collected to train and verify the desire inference model. Average precision was 81% for human desire inference.

The remainder of this dissertation is organized as follows: Chapter 2 briefly reviews related works. Chapter 3 describes the desire inference process that explains our strategy to exploit contexts variables from observations, and presents how desire can be inferred via the desire inference model. Chapter 4 depicts our extended computational framework by applying it into DBBNs. Chapter 5 presents our pilot study to identify and clarify relevant factors. Chapter 6 states data analysis using correlation. Chapter 7 shows the results of evaluation on the desire inference model. Finally, Chapter 8 concludes this dissertation including future works.

CHAPTER 2. RELATED WORKS

2.1 Defining Situation

Situ [6] presents a situation-theoretic approach to human-intention-driven service evolution in context-aware environments. In that paper, authors define situation that is rich in semantics and useful for modeling and reasoning human intentions. Accordingly, intentions are defined based on the observation of situations.

Situation at a time *t*, *Situation (t)*, is a triple $\{d, A, E\}_t$, in which *d* is the predicted user's desire, *A* is a set of actions for achieving *d*, and *E* is a set of context values with respect to a subset of the context variables at time *t*.

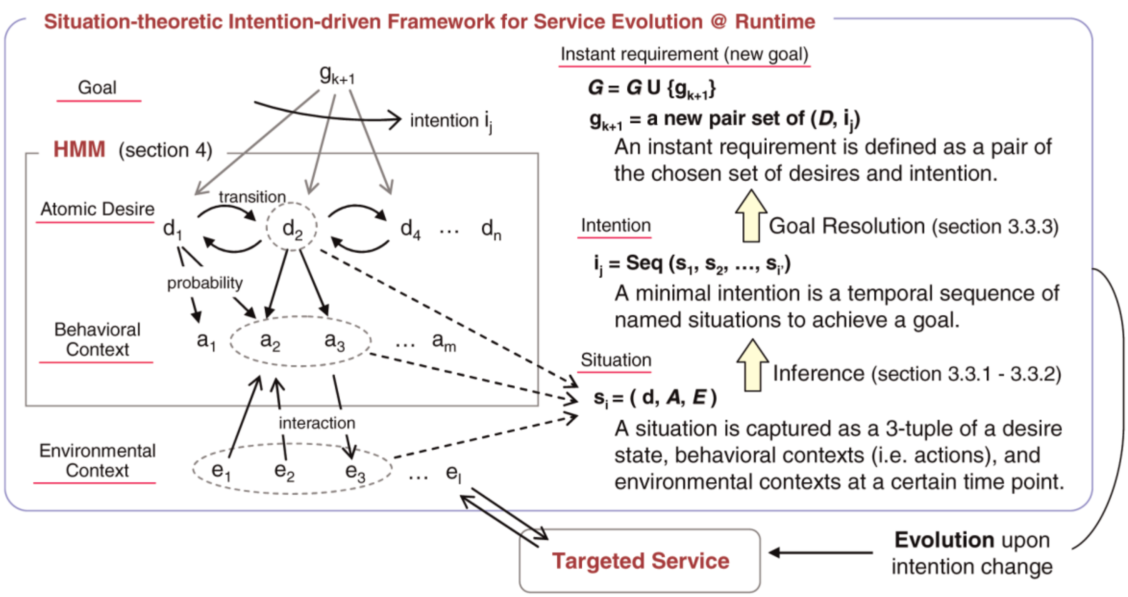


Figure 1. *Situ*: Framework for service definition and packaging with runtime software evolution.

By detecting the desire of individual as well as capturing the corresponding context values through observations, such a computational framework allows researchers to model and infer human intentions. An inference process based on Hidden Markov Model (HMM), as a suitable mathematical model to address the runtime prediction issue, makes instant definition of individualized services possible. Thus, situations perceived by a human depend on the human's internal mental state and the actions performed can be regarded as an external reflection.

Also, as situations [7] are semantic abstractions from low-level contextual cues, human knowledge and interpretation of the world must be integrated into a model of situation representation. This can either be done during a specification process, i.e., a human defines the situations and their relationships based on his knowledge, or situations are recognized and learned automatically, i.e., sensor perceptions are aggregated and associated to a human-defined situation label using machine learning techniques.

The latter relates to the domain of human activity recognition. Most approaches in this area focus on the classification of basic human activities or scenarios, without considering a richer contextual description.

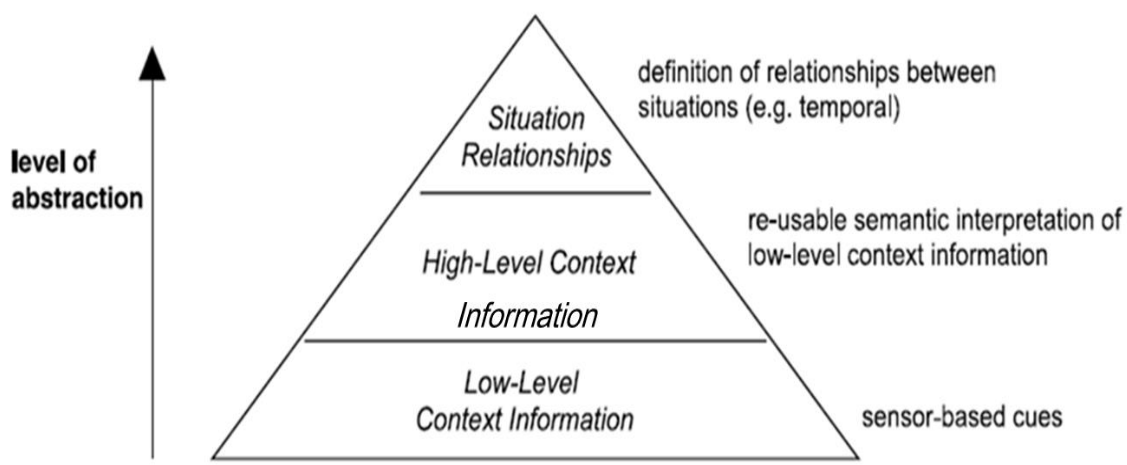


Figure 2. Overview of the different layers of semantic context interpretation and abstraction

2.2 Activity Recognition

In human activity recognition, much early work has been done in computer vision. They leverage on video camera, and explore various spatial-temporal analyses to recognize people's action from video sequences [8, 9, 10, 31, 41]. Recently, researchers are interested in recognizing activities base on sensor readings. For learning and recognizing human (inter)actions and behavior models from sensor data, many approaches have been proposed in recent years, with particular attention to applications in video surveillance, workplace tools, and group entertainment. Some projects focus on supplying appropriate system services to the users, while others focus on the correct classification of activities. Most of previous work in these activity recognition techniques is based on video, audio, or multimodal information using statistical models for learning and recognition. Recognition models are typically probabilistic based, and they can be categorized into static and temporal classification schemes. Typical static classifiers include naïve Bayes, decision tree, and k-nearest neighbor (k-NN). In temporal classification, state-space models are typically used to enable the inference of hidden states (i.e., activity labels) given the observations using Hidden Markov Model (HMM), Dynamic Bayesian Networks (DBNs) and Conditional Random Field (CRF) [9, 10]. The variants of CRF have been used to model complex activities of a single user. Most of the reported work has been concerned with the recognition of the activities of individuals who have been identified a priori.

2.3 Affective Computing

Currently, many researchers in the field of HCI are interested in user's emotional aspects because such affective states operate as indications of the user's internal mental state, desire and intention. The research on emotions in computer science is termed "Affective Computing", which is defined as "computing that relates to, arise from or deliberately influences emotions" [11, 13, 14, 42, 45].

Traditionally, this approach is divided into two major branches according to research interest. The first branch is the study of mechanisms to recognize human emotions or to express emotions by the machine used in the human-computer interaction system. The machine should interpret the emotional state of humans and adapt its behavior to them, giving an appropriate response for those emotions. Detecting emotional information begins with passive sensors which directly capture data from the user's physical state or behavior without interpreting the input. Most relevant works for emotion recognition have been focused on the low-level mapping between sensory data and underlying emotions. Recognizing emotional information requires the extraction of meaningful patterns from the gathered data. This is done by parsing the data through various processes such as facial recognition, gesture analysis and speech recognition.

The second branch is the investigation on the simulation of emotion or emotion synthesis with machines in order to discover more human emotions and to construct more realistic robots. As such, affective states including sadness, happiness, fear, anger, surprise and so on can be simulated. Especially, for emotion synthesis, there is the OCC (Ortony, Clore, & Collins) computational model that has 22 emotions categorized by valenced reaction to situations constructed either as being goal relevant events, as acts of an accountable agent, or as attractive or unattractive objects [5].

Also, human beings have abundant emotions, such as happiness, sadness, guilty, pride, shame, anxiety, fear, anger and so on. Affective state estimation uses pattern recognition, information retrieval methodology and EEG with brainwave [12] from the view of computational theory. For these methodologies, HMMs [13], Bayesian Networks (BNs), Fuzzy rules [14], and

Bayesian learning [15] are often employed. Most of these research efforts focus on the low level mapping between sensory data and underlying emotions. There are two main categories for these methodologies. First of all, researchers use sensory measures as predictors and apply classification algorithm 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 found with data. Second, researchers use the prior knowledge and expertise in graphical networks pertaining to using BNs and HMM models. 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.

Assessments of the user's affective state in terms of valence and arousal used Bayesian Networks (BNs) [16]. The facial and speech information of the user are considered as observable data. The temporal emotion-state structures are captured by a Dynamic Decision Network (DDN) model for a simple emotional state assessment task [17]. Also, a Dynamic Decision Network (DDN) model is applied to assess students' emotion in educational games [17]. The emotion states are modeled as consequence of how the current action and help fit with the student's goals and preferences. Somebody expressions are also used as evidences.

2.4 User Modeling

In traditional human factors and ergonomics research, the human operator is "fixed" once the initial configuration of the machine is done. Oftentimes there is only consideration for the goodness of "average" operators. However, human subjects may frequently enter abnormal or negative states, being inattentive, fatigued, or nervous.

On the other hand, a user model is the core component of an intelligent assistance system that captures, interprets and reacts to the goal, intention, need, and other internal states of the human user. As in the vision-augmented driving assistance systems [18], the traditional "external" assistance system is based on the assumption of normal operators, fully capable of fulfilling tasks. Such an assistance system aims to modify working conditions or raise alarms based solely on context awareness, e.g. detection of obstacles or departure from the lane center. The assistance is passive and very limited. To provide appropriate assistance we need pay more attention to human situations because it is in these states where the human user's operating performance deteriorates sharply; such operations tend to cause accidents, and thus the assistance is needed the most.

We require an underlying user model that fulfills critical tasks through the exchange of knowledge with decision and actuating components and under the requirement of high fidelity [1, 14, 19, 20]. A variety of user modeling tools have been developed, falling into two major categories: cognitive modeling from psychological and cognitive sciences, and statistical/probabilistic modeling based on various mathematical representations, such as rule-based systems, regression model, neural networks and BNs [21, 22]. Furthermore, user modeling has traditionally focused on what is generally considered "rational" aspects of user behavior, typically the user's knowledge and belief state. While useful, models focusing squarely on these aspects of user state often miss critical components of user mental state and behavior, in particular, the affective states. Generally speaking, "extra-rational" factors in human cognition and decision making, affective states, negative or positive, have been shown to strongly influence both reasoning and communication [23].

2.5 User-Need Inference and Assistance

Intelligent assistance systems need the ability to adaptively accommodate the user's specific need. In the READY system[24], the authors use Dynamic Bayesian Networks (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.

Extensive research applying Bayesian Networks (BNs) results in the creating intelligent software assistants. Among them, the Lumiere project in Microsoft Research is intended to help computer users with interactive interfaces [25]. By taking into account the user's background, actions and queries, DBN models are used to infer a user's goals and needs. 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.

CHAPTER 3. GENERAL DESIRE INFERENCE MODEL

3.1 System Overview

The goal of desire inference model is to accurately infer user's desire with context information. Human's emotion (i.e., a type of mental states) often leads to the motivational aspects of human behaviors. Thus, the desire inference model includes both user's behaviors and emotions, and depends on how to extract useful features from raw data. Figure 3 shows a schematic overview of the proposed desire inference system. The activity recognizer and emotion recognizer are practically applied to classify user's behaviors and emotions that a user tends to take based on the information collected from various multi-modal sensors, software, camcorders, and other objects suitably deployed in the Smart Home environments. Once sensor data is preprocessed, features can be successively extracted.

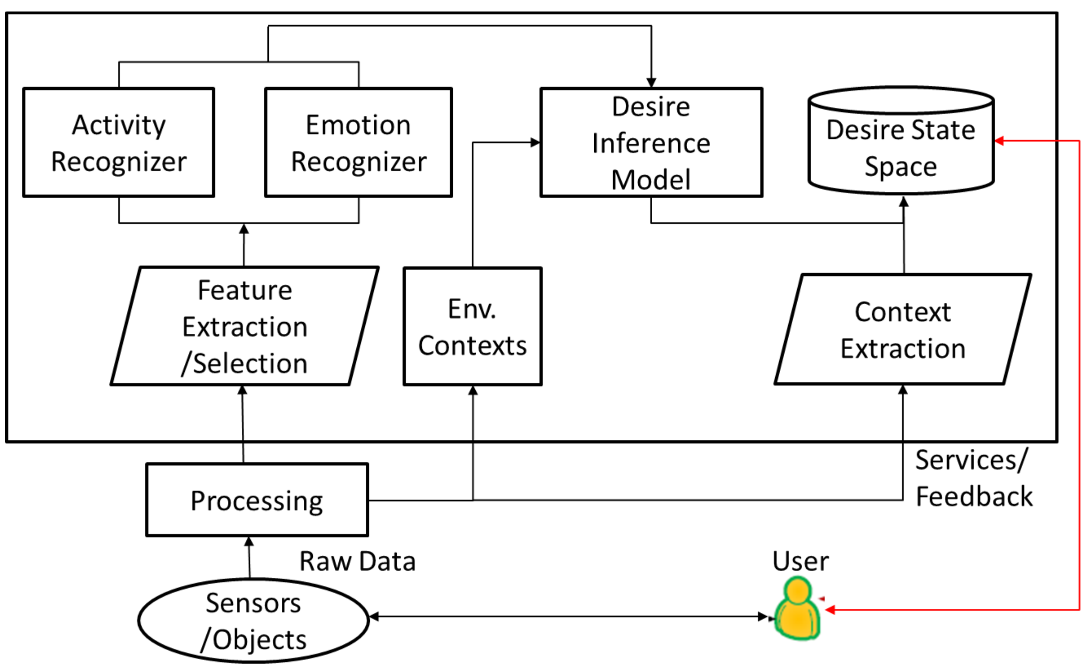


Figure 3. System Overview

The desire inference model perceives a user's interactions including the detailed information of different contexts (emotions, behaviors, and environment contexts), and then the most appropriate desire is subject to be inferred. All possible desires are defined as a desire state space. The system can deliver the service based on user's desire, and the users will provide the system with suitable feedbacks. In turn, the system, after analyzing the feedbacks, is highly expected to be updated with the new results. In the following section, we describe the details of each component.

3.2 A Description of Elements for Desire Inference Model

In this section, the overall elements of the desire inference process including environmental, behavioral and emotional contexts are described. Previously, we have defined situation in terms of behavioral and environmental contexts [6]. However, in order to reduce ambiguity in desire inference one may include the mental states M as one of the inputs together with both the actions A and the environmental contexts E. The perceived situation, or the sets M, A, E in which M is a set of the user's affective states, A is a set of the user's actions to achieve a goal, and E is a set of context values with respect to a subset of the context variables, is derived from raw sensor data first, and is then predicted by BBNs assuming that the human beliefs are known in advance. Figure 4 schematically shows the overall desire inference process.

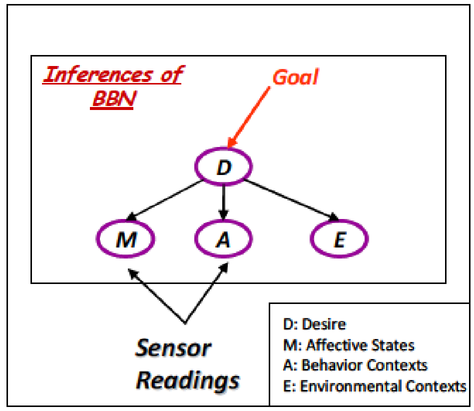


Figure 4. The Overall Desire Inference Process

A. Affective States

Human emotions provide emotional contexts (affective states), which result from sensory observations. The novelty of our approach is to apply affective states in distinct comparison with the previous studies. Affective states simply depict the results obtained through categorized types of emotion. An OCC model that specifies 22 types of emotion sorted by valenced reactions to situations is a well-recognized (or widely-used) computational model for emotion synthesis. Based on OCC model, the 7 basic emotions including happy, surprised, angry, disgust, fear, sad and neutral are selectively applied. Paul Ekman also determined that these basic emotions felt by all humans and the most basic emotions are universally recognized [45].

B. Behavioral Contexts

Behavior Contexts are empirically demonstrated as human actions. These context variables are derived from raw sensor data using different modalities including video recording, observation, and note taking.

C. Environmental Contexts

Environmental contexts categorized as lower level contexts are originated from sensing the entire environment surrounding a user being observed in a system. This type of contexts composed of time, location, and status of objects can be used to give rise to the corresponding actions. Especially, a set of environmental contexts are key factors to detect human emotions and actions, and other environmental contexts are related to the results created via interaction with emotions and actions.

3.3 Inference Model for Computational Desire

To answer the question what it is for an agent to have a reason to act, desire and belief are often considered as two major factors that stimulate us to act. Combining desire and belief efficiently provides a reasonable condition to some objective action.

In general, human desire is regarded as an inevitably motivating state due to its relation with internal motivation and is firmly determined by belief [36]. Even if we have a desire for something, and a belief that a certain action will bring up that thing, we may not get any explanation about the corresponding action unless we consider the belief and desire that caused that action.

In our study, the desire D is to be predicted by a decision making model, so called "BBNs," that deals with uncertainty, complexity, and probabilistic reasoning. On the basis of BBNs, we expect to put a specific design into a statistical inference process by considering the probabilities of events that can be observed, some evidences and her belief in the likelihood of other events.

A. Belief-Perceived Situation

In our work, we decompose contexts into emotional, behavior and environmental contexts applying to human belief in order to find the human desire. Based on this perspective, the perceived situation consists of human affective states, the actions that can be regarded as an external reflection and environmental contexts. Moreover, it is termed belief-perceived situation. Belief, one component of the Belief-Desire-Intention (BDI) model [52] encompassing belief, desire, and intention, is a function to express “internal core state” which can assume the causal-explanatory role. Belief defines the informational level to simultaneously deal with unchanged or very slowly changed information due to time interval necessary for modification of beliefs.

In this section, we formally define Belief-Perceived situation as follows.

Definition. A Belief-Perceived situation is B(m, a, e) for $m \in M$, $a \in A$, $e \in E$, where B is a Belief function, and the triple <M, A, E> in which M is a set of the user’s affective states, A is a set of the user’s actions and E is a set of context values with respect to a subset of the context variables.

Hypothesis. Desire can be derived from belief-perceived situations.

Belief-perceived situations are the formula representing key elements for desire inference, i.e. B referring to BBNs, M referring to affective states, A referring to behavior contexts and E referring to environmental contexts.

B. Desire Inference Model

BBNs are a decision-support framework for fixing the problems involving uncertainty, complexity, and probabilistic reasoning. BBNs are mainly used for situations that require statistical inference – in addition to statements about the probabilities (i.e., likelihood) of events, the user knows some evidence (i.e., some events that have actually been observed) and furthermore wishes to update her belief in the likelihood of other events, which have not as yet been observed.

We introduce a conceptual Desire-Inference-Model which is the BBN model as schematically illustrated in Figure 5. This model is used for inferring the user's desires from observations with different modalities. In this proposal, we suggest three modalities including facial and speech recognition, and gesture analysis to obtain affective states. As illustrated in Figure 5, such model captures desires, the user's actions, affective states, and environmental context information.

- Desire. This component represents the user's desire. It constitutes the hypothesis we want to infer.
- Actions. This component represents the user's behaviors.
- Affective States. This component represents the user's emotional states including happy, surprised, angry, disgust, fear, sad and neutral.
- Environmental Contexts. This component represents information about the specific environmental factors that can influence the user's actions and affective states as well as desire.
- Observations. This component consists of sensory observations in different modalities characterizing individual user behavior and emotion.

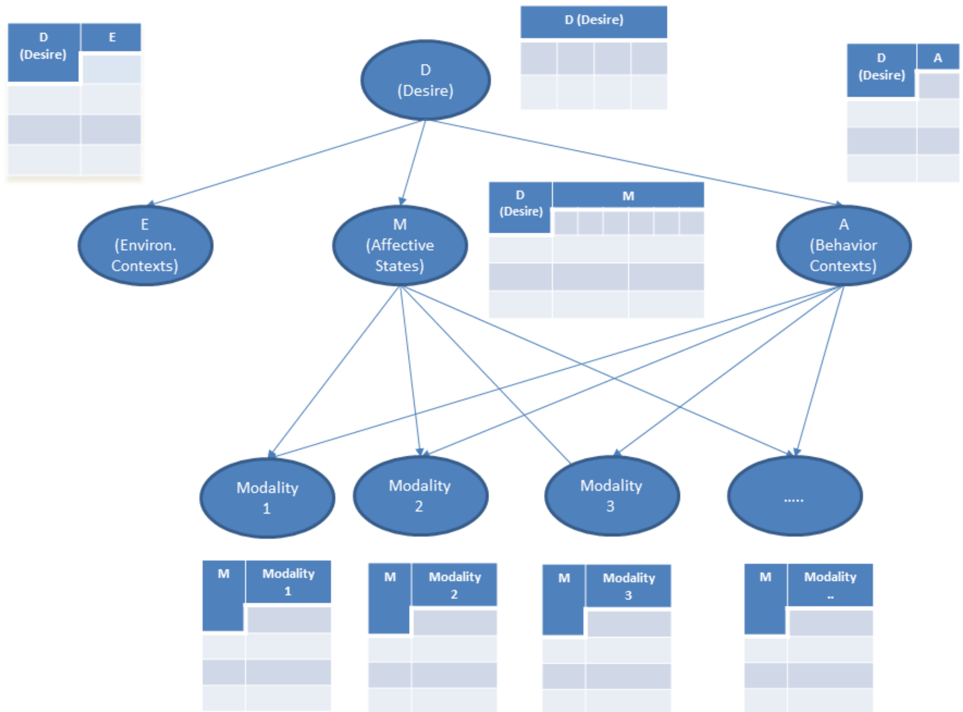


Figure 5. Conceptual Desire Inference Model

Belief, one of several factors for desire inference, is not included in this figure. In BBN, beliefs reflect a probability that a variable will be in a certain state based on the addition of evidence in a current situation. A-priori-beliefs are a special case of beliefs largely based on prior information. A-priori-beliefs are determined only by the information stored in the belief network's CPTs (Conditional Probability Tables). Moreover, evidence is information about a current situation. BBNs support vague, contradictory, and incomplete evidence by allowing one to enter a probability for the evidence of a variable being in each state. This model highlights the causal relationship between actions, affective states, environmental contexts and observation variables as represented by arrows. The environmental contexts variables influence the user's action and affective states as well as desire.

3.4 Computation of Desire by Static BBNs

The desire inference is to estimate user desire from the evidence of context variables collected by observations using a statistical inference technique. In this section, we introduce the notation for the inference process using the quantitative model, namely the CPTs.

A. Conditional Probability Tables (CPTs)

Once the topology of our desire inference model is specified, the next step is to compute the relationship between connected nodes, done by specifying a conditional probability distribution for each node. All nodes have CPTs. Conditional probabilities represent likelihoods based on priori information or past experience. For example, for the node "Desire", P(Desire) denotes the priori distribution of the variable "Desire".

In general, CPTs are obtained by statistically analyzing a huge amount of training data. For this research, the initial values of the CPTs for observations result from a following source. We refer to several large-scale subjective surveys to obtain initial CPTs [27]. Then, the overall initial data can be taken according to the Bayes inference rule. Table 1 shows one part of a hypothetical CPTs to illustrate our model. Then the existing learning algorithms can be used for training a model using the training data [28]. Furthermore, the initial CPTs are automatically refined to match each individual subject.

Table 1. One Part of CPTs in the Classroom Example

M	Facial Recognition		
	Eyebrows Up /Lip Down	Eyebrows Down /Lip Up (Smile)	Neutral
Happy	0.1	0.72	0.18
Surprised	0.48	0.35	0.17
Angry	0.5	0.32	0.18
Disgust	0.4	0.47	0.13
Fear	0.45	0.25	0.3
Sad	0.8	0.1	0.1
Neutral	0.03	0.15	0.82

Table 2 shows CPTs using our real data which is collected from pilot study in chapter 5. They shows the results between emotion and brainwave at a single time instance. Using CPTs, we can calculate conditional probability of a single time. Thus, we need to consider time series to infer human desire because user's desire is changing time by time.

Table 2. Results of CPTs between Emotion and Brainwave in Static Desire Inference Model

M	**Brainwave**			
	Theta	Alpha	Beta	Gamma
Happy	0.82	0.13	0.03	0.01
Surprised	0.13	0.08	0.68	0.11
Angry	0.18	0.21	0.58	0.12
Disgust	0.06	0.16	0.07	0.71
Fear	0.01	0.33	0.11	0.54
Sad	0.10	0.09	0.66	0.15
Neutral	0.43	0.48	0.06	0.03

B. Derivation of a Desire

BBNs-based human desire inference aims to estimate human desire from evidences, or context variables, collected by observations such as facial and speech recognition, and gesture analysis using a certain inference technique. We first introduce the notations. Under these notations, the desire inference model specifies two probabilistic relationships: the desire inference transition model P(D | M, A, E) and the evidence generation model P(M, A, E | D). An inferred desire through the highest conditional probability of P(D) computed from BBNs is shown on CPTs. Thus, the desire inference is derived from Bayes' rule as follows [29];

$$\mathrm{P(D|M,A,E)} = \frac{P(M,A,E|D)P(D)}{P(M,\ A,E)}$$

$$= \frac{P(M,A,E|D)P(D)}{\sum P(M,A,E|D_i)P(D_i)}$$

When applied, the probabilities of the Bayes' rule are directly used as a part of particular statistical inference results at a single time instant. With the Bayesian interpretation of probability, the Bayes' rule represents how a subjective degree of belief should rationally change to account for evidence: this is Bayesian Inference, which is fundamental to static Bayesian Networks.

CHAPTER 4. DYNAMIC DESIRE INFERENCE PROCESS

4.1 Extended to Dynamic Desire Inference Model

The BBNs is a probabilistic graphical model (i.e., a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). Formally, Bayesian networks are acyclic graphs whose nodes represent random variables including observable quantities, unknown parameters or hypotheses. Efficient algorithms are employed for performing inference and learning in Bayesian networks. Static BBNs can be work with evidences and beliefs at a single time instant. Thus, static BBNs are not enough in modeling systems that evolve over time [42, 43]. Bayesian networks that model sequences of variables are generally called Dynamic Bayesian Belief Networks (DBBNs). Generalizations of Bayesian networks that can represent and solve decision problems under uncertainty are called influence diagrams [33, 39]. The dynamic inference process based on DBBNs is being considered to allow the system to accommodate emerging new desires and adjust user's change in inclination towards certain desires over a period of time. DBBNs formalism is based on the Bayesian networks with extensions to represent discrete sequential systems.

Our central hypothesis is that DBBNs serve as an effective computational model for inferring user's desires when provided with real-time observations and finite historical data of the user's affective states, environmental contexts, and behavioural contexts as inputs. Based on our preliminary results, the problem of ambiguous human desire inference has been identified when human affective states are not considered. The feasibility of using DBBNs to model human desires has been established.

Thus, Dynamic Bayesian Belief Networks (DBBNs) have been developed to overcome the limitation which does not evolve over time. As a result, our study has extended the computational model from BBNs to DBBNs. In general, a DBBN is made up of interconnected time slices of static BBNs, and the relationships between two neighbouring time slices are modelled by Hidden Markov Model (HMM), i.e., random variables at time t are affected by the observable variables at time t, as well as by the corresponding random variables at time t-1.

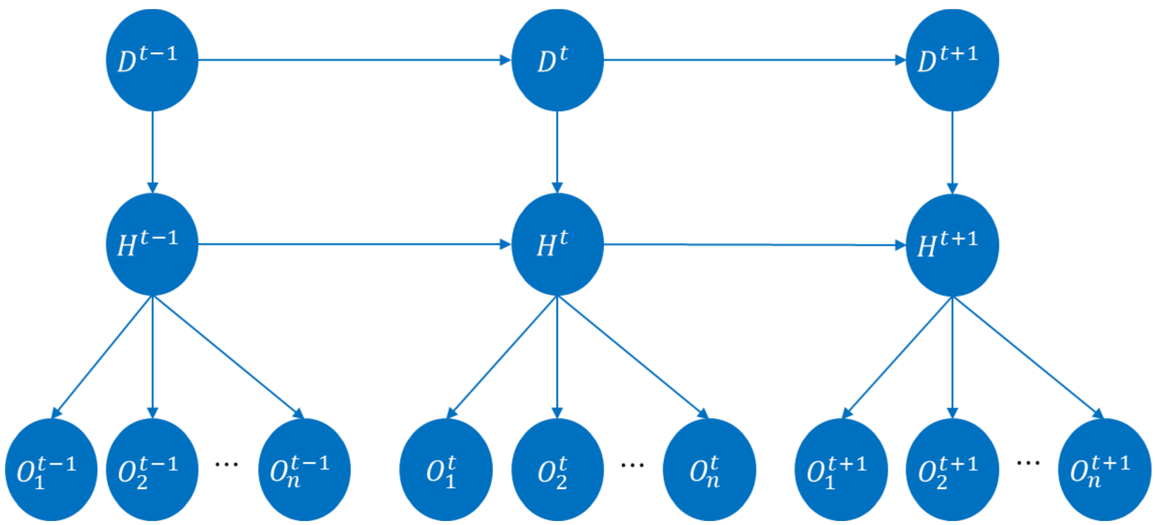


Figure 6. DBBNs consisting of three time slices, where O represents a collection of hypothesis nodes. H is a collection of hidden nodes, O is a collection of observation nodes, D and t indicate desires and time, respectively.

Figure 6 illustrates DBBNs. DBBNs represent a generalization of conventional systems for modeling dynamic events, such as Kalman filtering and HMMs. DBBNs provide a very powerful tool by providing a coherent and unified hierarchical probabilistic framework for sensory information representation, integration, and inference over time. Furthermore, DBBNs provide us with the ability to predict the influence of possible future desire through its temporal causality.

Our generic framework to apply BBNs to user modeling is the Desire Inference Model. It is used to infer user's desire from observations. The desire of user and the hidden nodes in current time slice are influenced by the corresponding variables in the most recent time slice. The hidden nodes are not observed at all, but are expected to interact with observed nodes. Each node has a probablity distribution over the possible output variables and then resulted in a sequence of output parameters. The resulting sequence ouput is subject to be calculated by an DBBNs [39].

4.2 Feature Selection

It is of great importance to collect and integrate information needed to infer human's desire in a timely manner. We collected the observations from informative sensors and observing software in order to infer human's desire efficiently and timely. For selecting more effective observations from a large set of observations obtained from deployed sensors so as to enable our learnt model to be more discriminative among various desires, information gain is calculated for each pair of desire and observation. The possible desires are formulated as $\mathrm{D} = \{d_1, d_2, \cdots, d_n\}$ and mutual information (or information gain) [32] can be calculated. Important measure of information is entropy, which is usually expressed by the average number of bits needed to store or communicate one symbol in a message. Entropy quantifies the uncertainty involved in predicting the value of random variable [56].

Thus, mathematically, the user's desire inference problem may be viewed as mutual information, with the entropy, H, of a discrete random variable D that is a measure of the amount of uncertainty associated with the value of D. It is zero when D is unambiguous, i.e., when one state has a probability of 1. A discrete random variable D, $\mathrm{D} = \{d_1, d_2, \cdots, d_n\}$, represents the set of possible desires. The sensory observation O is achieved form m diverse sensors and observing software, i.e., $\mathrm{O} = \{O_1, O_2, \cdots, O_m\}$. Each set of O includes sets of M, A, E. The goal is to estimate the posterior probability that D $= d_i$ is true given O, i.e., p(D $= d_i|O$). According to the information theory, the entropy H over the hypothesis variable D is calculated as follows:

$$\mathrm{H(D)} = -\sum_{i} p(d_i) \log p(d_i)$$

Mutual information measures the amount of information that can be obtained about one random variable by observing another. Given the beliefs in hypothesis for the last time slice D^{t-1}, the mutual information of a sensory observation O_j to current hypothesis D^t can be denoted as $\mathrm{I}\left(D^t;\ O_j\right)$.

$$\begin{aligned} \mathrm{I}\left(D^t;\ O_j\right) &= H(D^t) - H(D^t|O_j) \\ &= H(D^t) - \sum_j p(o_j) H(D^t|o_j) \\ &= -\sum_i p(\,d_i^t|D^{t-1}) \log p(d_i^t|D^{t-1}) \\ &\quad + \sum_j [p(o_j) \sum_i p(d_i^t|\,D^{t-1}, o_j) \times \log p(d_i^t|\,D^{t-1}, o_j) \end{aligned}$$

where

$$\mathrm{H(D^t)} = -\sum_i p(d_i^t) \log p(d_i^t)$$

$$\mathrm{H}\left(D^t|O_j\right) = -\sum_j p(o_j) \sum_i p\left(d_i^t|o_j\right) log p(d_i^t|o_j)$$

The above equation is fundamental for dynamically computing the uncertainty reducing potential for D due to O. We could extend it to consider the case that multiple sensors, $O = \{O_1, O_2, \cdots, O_n\} \subseteq \boldsymbol{O}$, are instantiated simultaneously

$$
\begin{aligned}
&\mathrm{I}(D^t; O_1, \cdots, O_n) \\
&= \sum_i p(d_i^t | D^{t-1}) \log p(d_i^t | D^{t-1}) \\
&\quad + \sum_{o_1} \cdots \sum_{o_n} [p(o_1, \cdots, o_n) \\
&\quad \times \sum_i p(d_i^t | D^{t-1}, o_1, \cdots, o_n) \\
&\quad \times \log\, p(d_i^t | D^{t-1}, o_1, \cdots, o_n)]
\end{aligned}
$$

In the above equation, the probabilities are readily available from the forward and backward inference propagation based on hypothesis beliefs for last time slice. $\mathrm{p}(d_i^t | D^{t-1}, o_1, \cdots, o_n)$ is the posterior probability of hypothesis state for current time slice given a configuration on sensor state and the beliefs in hypothesis in the last time slice. $\mathrm{p}(d_i^t | D^{t-1})$ is the posterior probability of hypothesis state without acquiring new sensory observation.

After estimating the information gain of each pair of desires and sensory observation, the sensory observation selection chooses those sensory observations with higher mutual information. These selected observations are more highly associated to their corresponding desires. In sensor selection, $\mathrm{H}(D^t)$ has the same value for all sensors and need not be calculated.

4.3 Learning

A. Different levels of creating DBNs

Creating DBNs from data can be characterized as a very complex problem. There are four cases to describe these problems, as shown it Table 3 [33]. Full observability (complete data) means that the values of all variables are known. Partial observability means that we do not know the values of some variable. Such case exists because in some situations variables cannot be measured, and then they are called hidden variables. It is possible that they can be measured in training data, but they are not, and then they are termed as missing variables. Unknown structure means that we are not in position to know the whole topology of the network.

Table 3. Methods for creating DBNs structure and determining their parameters [33, 34]

Structure/Observability	**Method**
Known/Full (Complete data)	Simple statistics
Known/Partial (Incomplete data)	Expectation Maximization (EM) algorithm or Gradient Ascent algorithm
Unknown/Full	Search through model space
Unknown/Partial	Structural Expectation Maximization (SEM) algorithm

In our computational model, we have a number and type of nodes in the network, but we do not have the knowledge of their relation and mutual independence. Thus, we should be interested in finding the way to learn the structure of DBN from observable data and expert knowledge about the domain.

Algorithms that deal with such problems can be grouped into two categories [34]. One category of algorithms uses heuristic searching methods to construct a model and then evaluates it using a scoring method. The other category of algorithms constructs Bayesian networks by analyzing dependency relationships among nodes. The dependency relationships are measured using one kind of several Conditional Independence (CI) test types.

B. Learning (Training)

To take into account temporal information and relationship between an desire and its corresponding sensory observation, we used Dynamic Bayesian Belief Networks (DBBNs), which models time information and predicts probability of an activity. Figure 6 shows the graphical structure of our proposed human desire inference model.

For each time slice t, the possible desire is defined as $\mathrm{D}^t = \{d_1^t, d_2^t, \cdots, d_n^t\}$ and the sensory observation extracted at time t is denoted as $O^t = \{O_1^t, O_2^t, \cdots, O_m^t\}$. Hence, the problem to predict the desire given the previous desire estimates and the observation at t can be expressed as $\mathrm{p}(D^t|D^{t-1}, O^t)$. The parameters of an desire model are trained with Expectation Maximization (EM) algorithm [35]. The EM algorithm is used to find the maximum likelihood estimate (MLE) of the marginal likelihood by iteratively applying the two steps such as Expectation step (E Step) and Maximization step (M step). E step is to calcuate the expected value of the log likeli hood function, and M step is to find the parameter that maximizes this quantity. In desire inference model, we represented $\mathrm{P}(O^t|\theta)$ as E step and $\theta^* = argmax_\theta\, \mathrm{P}(O^t|\theta)$ as M step.

The desire inference model estimates parameters by evaluating $\theta^* = argmax_\theta\, \mathrm{P}(O^t|\theta)$ where θ is the set of parameters of the desire inference model and O^t is the set of the sensory observation. After training the desire inference model, posterior probability can be represented as $\mathrm{P}(D^t|O^t)$ based on a Bayes Filter.

4.4 Relationship between Contexts and Desire

The desire inference model is constructed with the information of contexts. With the different level of the expression power of the context, we can regard the current contexts, which are recognized by M, A, E as high level context data and utilize them as the input of a desire inference model.

As for construction of a desire inference model, the underlying sensory data will be interpreted as lower level contexts via context interpreters in the Context Extraction unit based on domain knowledge. Among them, we selected highly correlated sensory contexts (observation) and theses contexts are formulated as a context vector $\mathrm{C} =< C_1, C_2, \cdots, C_m >$. The meaning of a context refers to some information related to an object of interest and its status.

After extracting sensory observation from diverse sensory data and software, we selected contexts to represent inputs of desire inference model by calculating information gain for each desire and contexts information. Let the desire that the user wants to do be formulated as $\mathrm{D} = \{D_1, D_2, \cdots, D_m\}$. Considering the temporal information and the desire our system may infer, a desire at time t is represented as $D^t \in \{D_1^t, D_2^t, \cdots, D_m^t\}$. Given the current selected contexts $C^t =< C_1^t, C_2^t, \cdots, C_m^t >$ and the state of the previous desire, the desire at time t can be inferred from $\mathrm{p}(D^t|D^{t-1}, C^t)$. The desire inference model also models the structure among desire and context information by usin DBBNs, and its parameters of desire inference model are trained with EM algorithm [35]. The inference of the desire inferred at time t can also be ahcieved by applying Bayes Filter.

To model the relation of each desire-contexts information, the mutual information-like weight function is calculated, and the equation is formulated as below:

$$\mathrm{W}(D_i, C_i) = \sum_i P(D_i\,, C_i)(log\frac{P(D_i,\ C_i)}{P(D_i)P(C_i)})$$

To model the relation of between desire and contexts information, there is a rank table constructed for each desire inference model which contains the information of weight function of contexts information. The weight vector of desire D_i can be denoted as $R_i = \{r_i^1, r_i^2, \cdots, r_i^n\}$ where $r_i^j = W(D_i, C_i)$. The contexts information whose weight function value is higher is more related to this desire.

CHAPTER 5. A PILOT STUDY ON DESIRE INFERENCE

5.1 Objective of the Pilot Study

Generating large datasets is a core research priority within Human Computer Interaction (HCI) research domain. To date, several open source datasets have been produced and shared with the research community; however, challenge still remains in providing sufficiently differing datasets with a completely accurate gold standard.

The objectives of the pilot study are 1) to collect quantitative data and context information suitable for inducing-desire, 2) to establish the correlation between environmental contexts, observed user behaviors, identified user emotions and the changing desire of the participants based on the data, 3) to identify the relevant factors for desire inference and 4) to train and validate the initial computational model using the data.

We consider the observable factors, or dynamic factors (not static factors). To identify and verify key observable factors for effective and accurate user desire inference, we develop and evaluate a computational model. It infers users' desires using observable data, both within the system as well as from the surrounding environments such as smart home environments. We establish the correlation between the potential key factors and human desire. For these objectives, we conduct the experimentation in controlled environments in our Smart Home Lab (SHL). In the experimentation, we controlled introduction and removal of stimulus to induce participant's desire. Desires are not forced but induced. Controlled and limited availability of stimulus allows us to create a dynamic Bayesian model.

In this pilot study, we assume several hypotheses. By proving hypotheses, we can reach a valid conclusion (i.e., desire inference) through identifying relationships among three parameters (i.e., M, A, E) and sub-parameters. It refers to the conceptual framework within which the experiment is conducted. Hypotheses are as follows:

1) Desire inference can be derived from three types of parameters (based on BBNs and DBBNs).

2) Three types of parameters have sub parameters.

3) Affective states can be derived from three types of recognition parameters (Facial, Speech & Gesture Analysis).

4) Affective states can help derive desire.

5) Behavior can be affected to infer the desire.

6) Some factors in the environmental contexts can be affected to infer the desire.

5.2 Setup of Experimentation

It is difficult to collect sensor data for users in a real home and it is a timing-consuming task. Aiming for a realistic data collection, we conduct a study in SHL among 24 university students. The SHL has been configured, for the pilot study, to resemble a studio apartment, which provides isolation from outside distractions. The entire study is conducted in SHL which is a sensor-rich environment and is also observed by all participants in a controlled environment. Within SHL, there are multiple computers with various sensors and appliances. The coordinator of the study monitors digital recordings of sessions from the SHL control room using multiple camcorders and microphones, while observing and annotating the participant's actions, gestures, and speech patterns.

Figure 7 is the overview of our sensor-rich environment in the Smart Home Lab at Iowa State University.

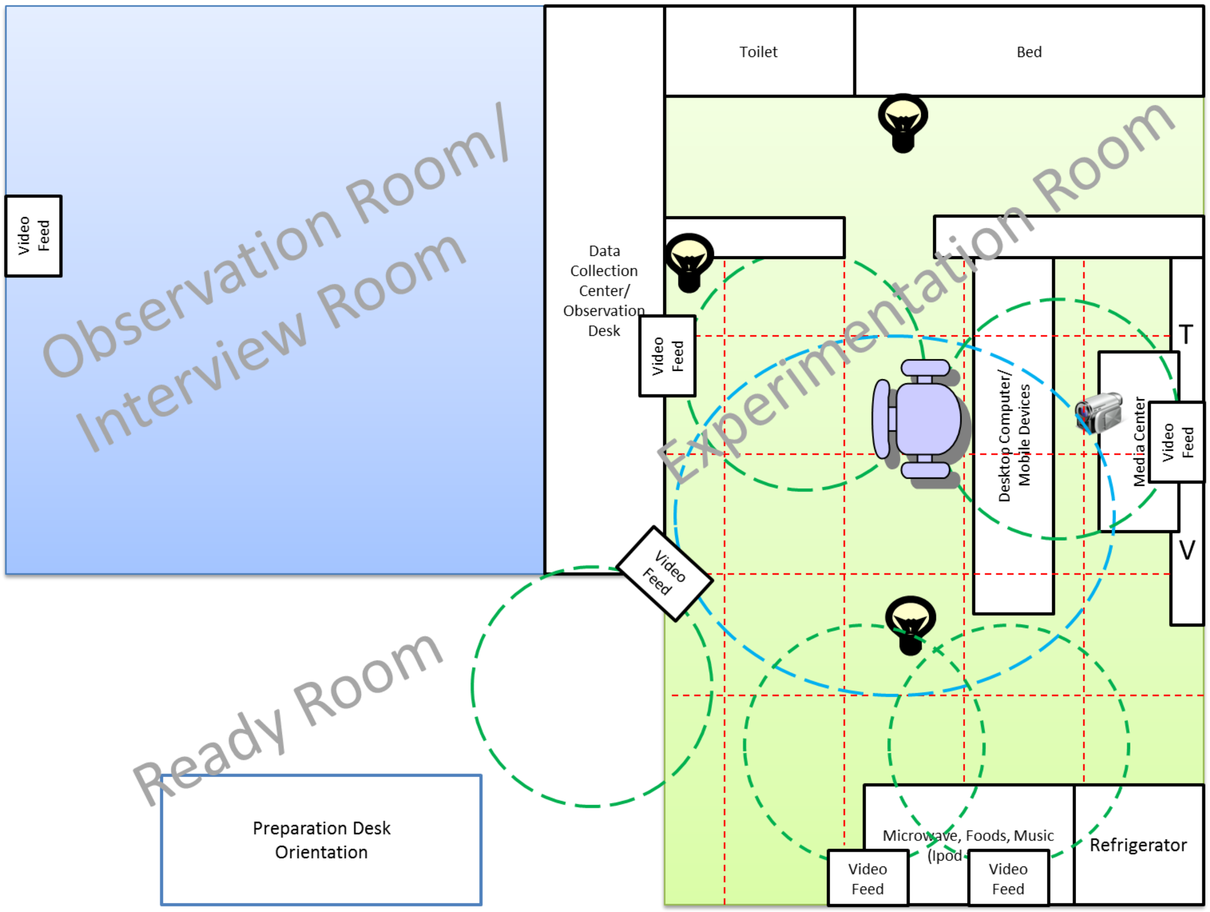


Figure 7. Overview of the experiment environment in Smart Home Lab

at Iowa State University

In the SHL, several sensors were deployed on different objects, and the location of each object is illustrated in Figures 7 and 8. During each session, about 1 hour of sensory data were recorded to validate our approach. The dataset was collected through 3 weeks by 24 volunteers. The recruitment criteria considered only adult participants, 18 years or older, who are physically & cognitively capable of performing some activities such as operating computers. Since the study involved watching videos, playing games, listening to music and seeing photos/comics, severely vision-impaired participants are not suitable for this study.

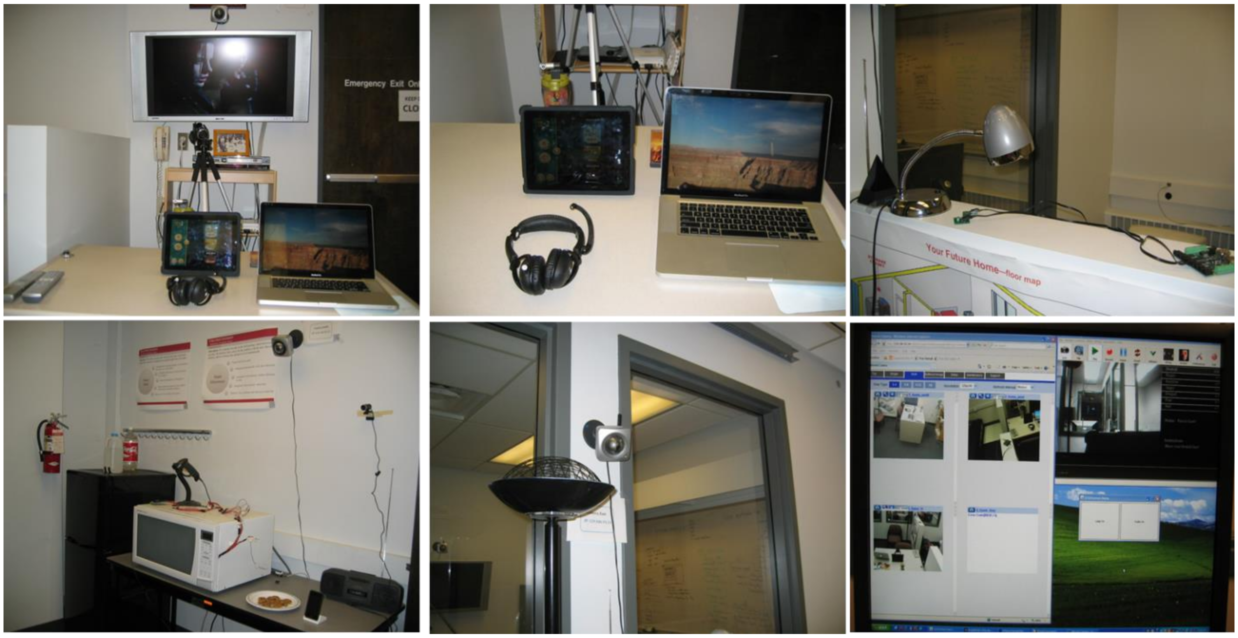


Figure 8. Setup of Experimentation in SHL. Turning on the lamp was used to synchronize the starting point when the study was started.

In Figure 9, we list all activities with their configurable activities states which are designed into our experimental environment. A set of annotation terms can be defined in a configuration to indicate the possible states. For example, a movie could be in either turning on, watching, change, pause, or turning off state.

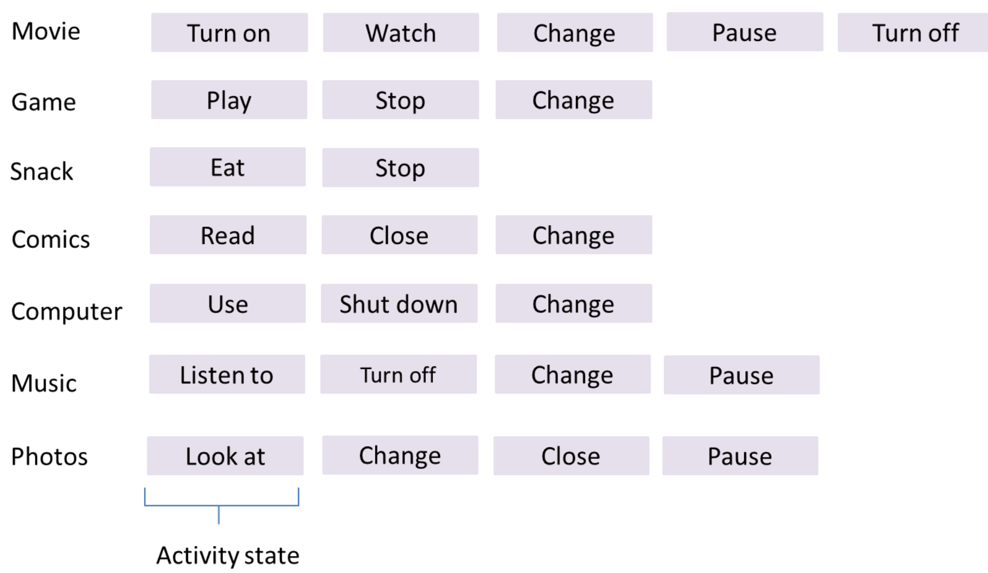


Figure 9. Configurable Activities States

A. Experimentation Procedures

During each session, the participants are told that they are free to do whatever they choose do within the SHL. However, they should explicitly identify the activity “they would like to do now”. Throughout the course of the study, additional stimuli typical to student’s life in their own dorm room, such as movies, comic books and novels, video game consoles, food and drink, were introduced and removed every 5 minutes to induce changes in participant’s desires and behaviors. At the end, each participant was asked to fill out a questionnaire, and go over the audio and video recording to provide additional comments and insights into their thought process and mental status. The information was annotated to the recordings, and any discrepancy will be identified and rectified. The environmental contexts were recorded via the existing SHL software. User’s behaviors were recognized both by behavioral recognition software as well as by human inspection.

The above study proceeded according to the procedures described as follows:

1. We scheduled a one-hour session with each participant. Each participant was required to come ten minutes before for orientation.

2. Once the participant arrived at the Smart Home Lab in Atanasoff Hall, we gave a brief introduction about the study to the participant. The participants took the direction and instruction on the procedure of the study as described here, as well as their rights including voluntary participation, no-fault policy for termination of participation in the study, and the effort to preserve the participant's privacy. Also, we surveyed allergy to ingredient of cookies (snacks), one of the stimuli applied in our study, because we wanted to protect participants first. If some participants had trouble and allergy, we would change the cookies (snacks) to suitable substitutes.

3. The participants were asked to name whatever they are looking at, doing or feeling as they go about their activities. This THINK-ALOUD method enables the researchers to see first-hand the process of activity completion. The researchers were asked to objectively take notes of everything that the participants say without attempting to interpret their actions and words.

4. Once the participant indicated that he/she is ready to proceed, the experiment was started by watching a movie. Also, this was recorded using web camera or camcorder and the participant was using the Bluetooth microphone. The participant's action was monitored and the researcher made a note at specific points such as action changes and emotion changes during the study.

5. Every 5 minutes, the participant was exposed to a different stimulus. Together with movies, 6 additional stimuli (game console, photo/pictures, cookies, audio system, comics, and computer) were aroused in sight.

During the experiments, we used smart appliances, sensors and observing software such as video camera, temperature sensor, mouse clicker and emotion recognition software in the SHL for capturing the activities that the participants perform. Two video cameras were set up in the SHL. A temperature sensor was also attached to the mouse. The mouse clicker and emotion recognition software were installed into the computer. When participants clicked the mouse button and used the computer, their mouse clicking activities were monitored and captured into the computer and video camera.

6. At first, the participant can start watching a movie. If they like to watch a movie, they could keep watching a movie. Otherwise, they could change to a different movie. After 5 minutes, when a new stimulus was provided, participants would do an activity using this stimulus if they want to use it. Otherwise, they will remain with the same activity like watching a movie. After using 7 stimuli, the participants could complete all activities by simply acting what they want to do with the stimuli.

For example, if they started watching a funny movie and they did not like it, they could change to different style. After 5 minutes, if we provided cookies as a new stimulus, they could change it again. If they did not like to eat cookies, they could ignore this stimulus and keep watching the movie. Every 5 minutes, when we provided a new stimulus, the participants could change their activities or they can keep doing a previous action. The staff would make a note when the activity of the participant changes in accordance to the defined activities or some unexpected changes.

7. The experiments were timed for 35 minutes. Descriptions on how various instruments were used for the study, and how we incorporated privacy protection mechanisms for their use, are described below.

Once data collection was completed, the participants were asked to fill out a questionnaire. We asked them to provide information regarding the activities they did, such as their desire when they changed from one activity to the other activity. The researcher requested the reason why they changed the activity and the emotion when their activities were changed. Following the questionnaire, our researcher interviewed and then reviewed taped video in order to discuss some specific points logged by the researcher during the experiment.

B. Methodology for Experimentation

In our study, the activities that the participant completed were designed to measure the metrics below. We used the following methodology to collect data.

1. Facial Recognition – We recorded facial movement as the one of the factors to get their emotion.
2. Speech Recognition – We recorded the voice of the participants as one of the factors to get their emotion.
3. Gesture Analysis – We recorded the gesture of the participants as the one of the parameters to get their behavior.
4. Environments Contexts – We considered the environment change as the one of the parameters.

Particularly, we used the following methodology to collect the data:

1. Video Taping - Emotion is one important factor to understand desire. So the video camcorder and web camera were used for recording during the study. The video files were annotated with the participant ID for protecting personal privacy.
2. Observation - One of our research personnel was on site during the study and serve as facilitator and observer. This person wrote down some activities of interest and annotates additional information that might be useful such as the change of activities.

3. Think-aloud - A method used to gather data in usability testing in product design and development. Think-aloud protocols were introduced to involve participants to both think and announce aloud as they were performing a set of specified tasks. Users were asked to say whatever they were looking at, thinking, doing, and feeling, as they went about their task. This enabled observers to see first-hand the process of task completion (rather than only its final product). Observers participating in such a test were asked to objectively take notes of everything that users announced, without attempting to interpret their actions and words. Test sessions were often audio and video taped so that developers could go back and refer to what participants did, and how they reacted. The purpose of this method was to make explicit what was implicitly present in the subjects who were able to perform a specific task.
4. Questionnaire - An ID was assigned to each participant at the time when they filled out the questionnaire.
5. Post-experiment Interview and review of taped videos - After finishing experimentation, we interviewed the participants. We wanted to make sure that their desire changes and our understanding of them was accurate. (We got the agreement of this process.)
6. Emotion Analysis Software - To understand emotion, we used the emotion analysis software such as that for facial emotion recognition and speech recognition.
7. Keyboard/Mouse Activities Logger - We used the logger software installed in the computer to track the details of the participants' actions during the study. Logger helped capture the activities of the keyboard and mouse performed on the computer screen by a user and stored them in a log file to be processed later.

CHAPTER 6. DATA ANALYSIS FOR RELEVANT FACTORS

We collected the data using the seven instruments: videotaping, observation, think-aloud, questionnaire, post-experiment Interview and review of taped videos, emotion analysis software and keyboard/mouse activities logger. Thus we have a dataset generated for each participant during her study session.

The actions of participants were monitored. The researcher made a note about specific points during the study. The researcher analyzed these specific points in post-experiment interview and review of taped videos. We have planned to do a correlation analysis also between different factors and desires. We create a correlation matrix between each factor and desire, and analyze the impact of that factor on desire; we do so for all other factors. If the factors and desire are almost highly correlated (positive/negative), it will help us to reduce the number of factors. This allows us to consider different models without favoring any particular formulation. Also we can determine the most effective model.

6.1 Data Coding

For analysis, data needs to be coded to an analytical process in which data, in both quantitative form (such as questionnaires results) and qualitative (such as interview transcripts), are categorized to facilitate analysis [37]. Data coding means the transformation of data into a form understandable by computer software. The classification of information is an important step in preparation of data for computer processing with statistical software. One code should apply to only one category and categories should be comprehensive. There should be clear guidelines for coders (individuals who do the coding) so that code is consistent. Some studies will employ multiple coders working independently on the same date. This minimizes the chance of errors from coding and increases the reliability of data.

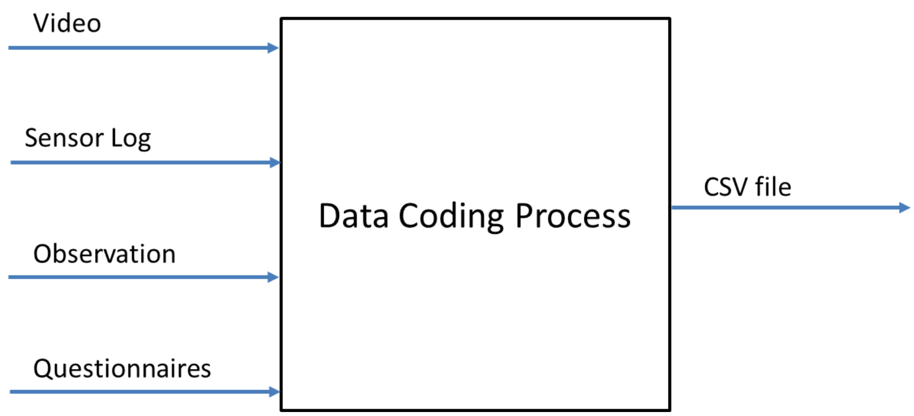


Figure 10. Process of Data Coding

In our study, we processed data coding scheme from raw data to CSV data format for statistical tool R [50]. As a result, we create 30 input formats which include 3 kinds of emotions (facial recognition, speech and brainwave), 24 configurable activities and 3 environmental contexts (location, temperature and light). Figure 10 shows the process of data coding and Figure 11 is an example of data coding.

```
Emotion   Recognition    Classification  Dump

First two  numbers   are  the  frame "number,"  and  elapsed
     time  in  Milliseconds
The  next  7   numbers   are  the  classification  "results,"

corresponding  to  "Neutral," "Happy,"  "Surprised,"
     "Angry,"  "Disgust," Fear  and  Sad  respectively.
0    0     0.999 0     0     0     0     0     0
1    453   0.995 0     0     0.003 0     0     0.002
2    922   0.004 0.001 0     0.788 0.069 0     0.138
3    1407  0.378 0     0     0.543 0.013 0     0.065
4    1860  0     0     0     0.885 0     0     0.115
5    2313  0     0     0     0.892 0     0     0.108
6    2766  0.001 0     0     0.809 0     0     0.189
7    3235  0.002 0.001 0     0.848 0.001 0     0.149
8    3672  0.045 0     0     0.86  0.004 0     0.091
9    4141  0.108 0.002 0     0.744 0.014 0     0.131
10   4625  0.126 0     0     0.727 0.035 0     0.112
11   5078  0.02  0     0     0.885 0.011 0     0.084
12   5532  0.128 0     0     0.76  0.007 0     0.105
13   5985  0     0     0.001 0.703 0     0     0.296
14   6469  0.005 0     0     0.9   0.002 0     0.093
15   6922  0.002 0.001 0     0.819 0.013 0     0.165
16   7375  0.209 0.001 0     0.637 0.011 0     0.142
17   7844  0     0     0     0.925 0     0     0.075
18   8313  0     0     0     0.894 0     0     0.106
19   8782  0     0     0     0.039 0.001 0     0.96
```

Emotion	Recognitio	Classificat	Dump								
First	two	numbers	are	the	frame	number,	and	elapsed	time	in	Milliseconds
The	next	7	numbers	are	the	classificat	results,				
correspon	to	Neutral,	Happy,	Surprised,	Angry,	Disgust,	Fear	and	Sad	respectively.	
0	0	0.999	0	0	0	0	0	0			
1	453	0.995	0	0	0.003	0	0	0.002			
2	922	0.004	0.001	0	0.788	0.069	0	0.138			
3	1407	0.378	0	0	0.543	0.013	0	0.065			
4	1860	0	0	0	0.885	0	0	0.115			
5	2313	0	0	0	0.892	0	0	0.108			
6	2766	0.001	0	0	0.809	0	0	0.189			
7	3235	0.002	0.001	0	0.848	0.001	0	0.149			
8	3672	0.045	0	0	0.86	0.004	0	0.091			
9	4141	0.108	0.002	0	0.744	0.014	0	0.131			
10	4625	0.126	0	0	0.727	0.035	0	0.112			
11	5078	0.02	0	0	0.885	0.011	0	0.084			
12	5532	0.128	0	0	0.76	0.007	0	0.105			
13	5985	0	0	0.001	0.703	0	0	0.296			
14	6469	0.005	0	0	0.9	0.002	0	0.093			
15	6922	0.002	0.001	0	0.819	0.013	0	0.165			
16	7375	0.209	0.001	0	0.637	0.011	0	0.142			
17	7844	0	0	0	0.925	0	0	0.075			
18	8313	0	0	0	0.894	0	0	0.106			
19	8782	0	0	0	0.039	0.001	0	0.96			
20	9235	0.002	0.001	0	0.861	0.001	0.001	0.134			
21	9704	0.95	0	0	0.038	0.001	0	0.011			
22	10125	0.461	0	0	0.497	0.016	0	0.027			

Figure 11. (a) Raw Data (b) Log file (C) CSV file

6.2 Statistical Analysis

In the experimentation, 24 volunteers participated in a one-hour session including pre and post session. During the study, real-time emotion software was used to collect emotion data. Through the camera, the emotion data was captured in real-time. After reviewing some of the data collected, it was discovered that the parameters we were trying to measure could not be ascertained from some participants due to them wearing hats or lowering their heads. Their faces were out of range of the camera, and emotion capture failed. Finally, among them, we used data of only 21 participants.

Normally, most existing studies done by other researchers used data during several days. Most of time, only one or two participants are involved in their study. Oftentimes no more than ten people participated in their pilot study. Nonetheless they still generated large data sets. In our case, 21 participants also provided very large data sets as the result of data collection with larger diversity of participants.

Each experimentation was executed in 35 minutes. When the moving picture is displayed, each frame is flashed on a screen for a short time (nowadays, usually 1/24, 1/25 or 1/30 of a second) and then immediately replaced by the next one. In North America and Japan, 30 frames per second (fps) is the broadcast standard, with 24 frames per second (24 frames/s) now common in production for high-definition video. In much of the rest of the world, 25 frames per second are standard. Thus, we used 24 frames per second and 1 person has 50,400 frames for each session. For preparation of data, totally, we consider 1,058,400 frames for behavior, 838,923 frames for emotions and 1,058,400 frames for environments.

A. Correlation

For effective and accurate user desire inference, it is important to identify and verify key observable factors. In this thesis, we established the correlation between the potential key factors and human desire. Correlation is any of a broad class of statistical relationships involving dependence which is any statistical relationship between two random variables or two sets of data [38].

Normally, there are 3 different kinds of correlation analysis that are Pearson, Spearman and Kendall. Pearson is the most familiar measure of dependence between two quantities. By dividing the covariance of the two variables, it is obtained by the product of their standard deviations. Thus, we adopted Pearson method, using the cor. test function in R.

Results of Correlation are between -1 and 1. If a result is -1 or 1, there is a perfect negative or positive correlation between the two values at all. On the other hand, if a result is 0, there is no linear relationship between two variables. It will be very rarely to get a correlation of 0, -1 or 1 [38, 44].

- High correlation : 0.5 to 1.0 or -0.5 to -1.0
- Medium correlation : 0.3 to 0.5 or -0.3 to -0.5
- Low correlation : 0.1 to 0.3 or -0.1 to -0.3

In Table 4, each desire is numbered.

Table 4. Defining Desire

	Desire
10	Want to watch a movie
20	Want to play a game
30	Want to eat cookies
40	Want to use a computer
50	Want to watch photos
60	Want to read comics
70	Want to listen music
80	Wild Cards

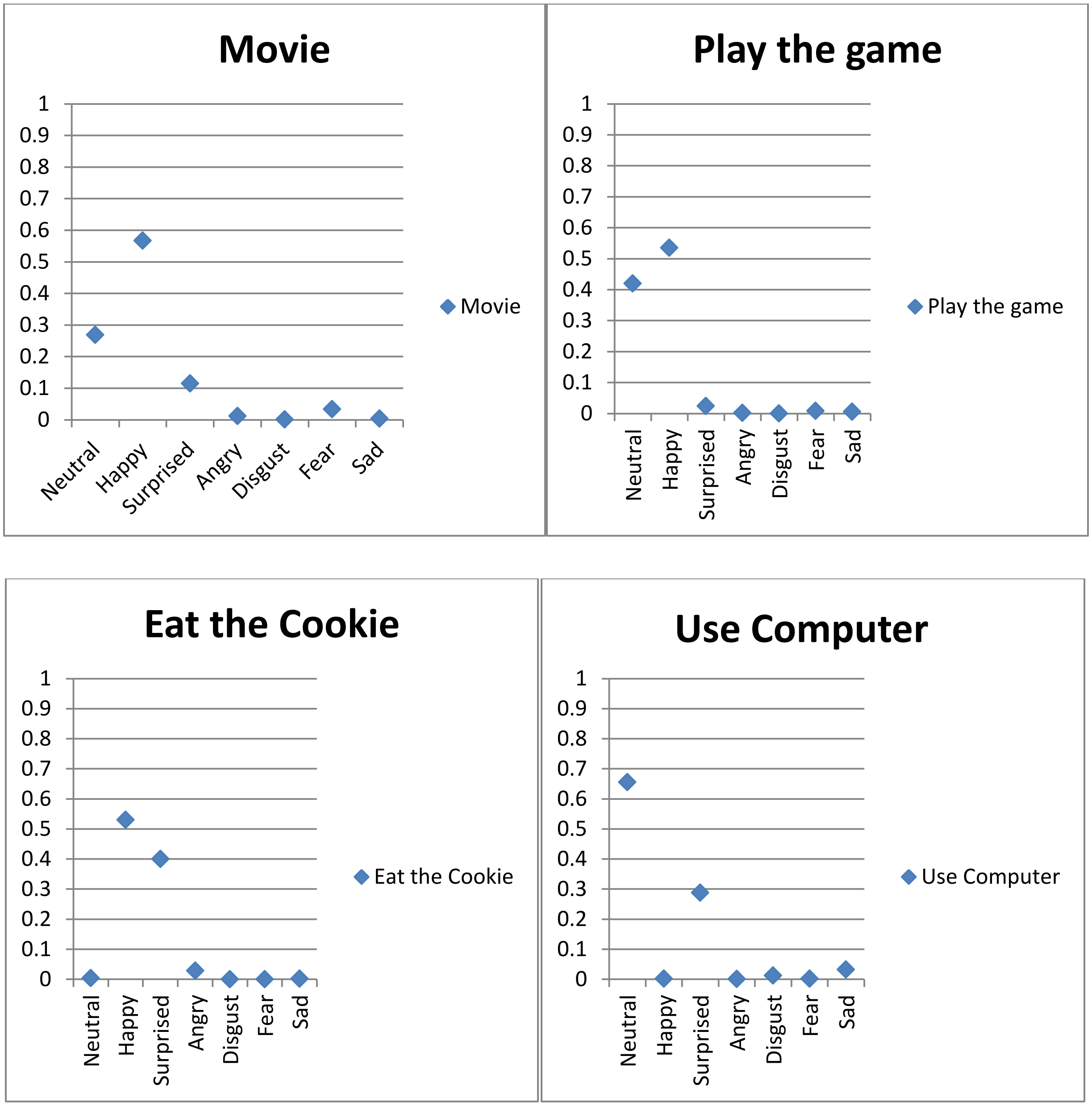


Figure 12. Each Desire-Emotion Correlation

Figure 12 Continued

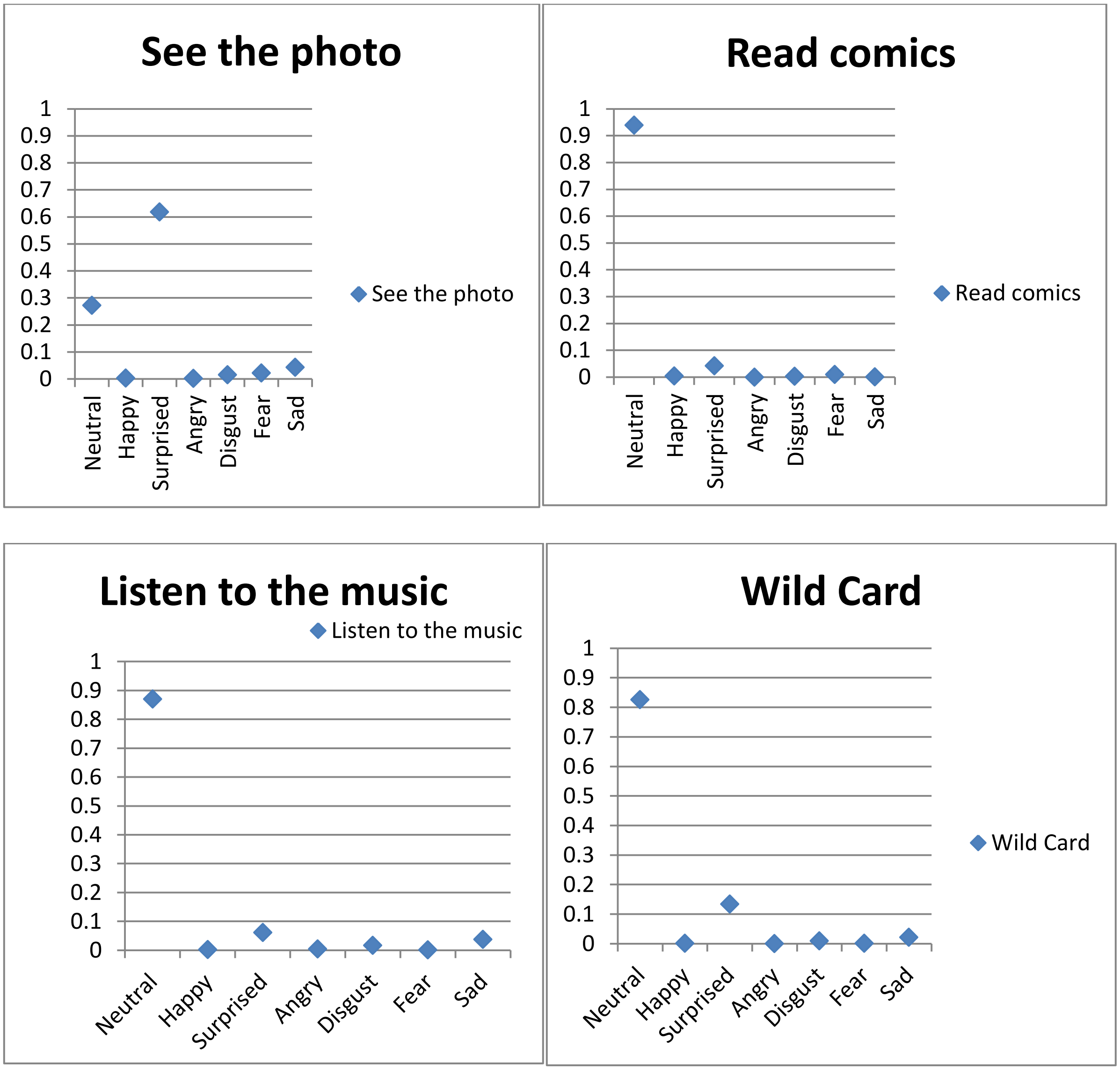
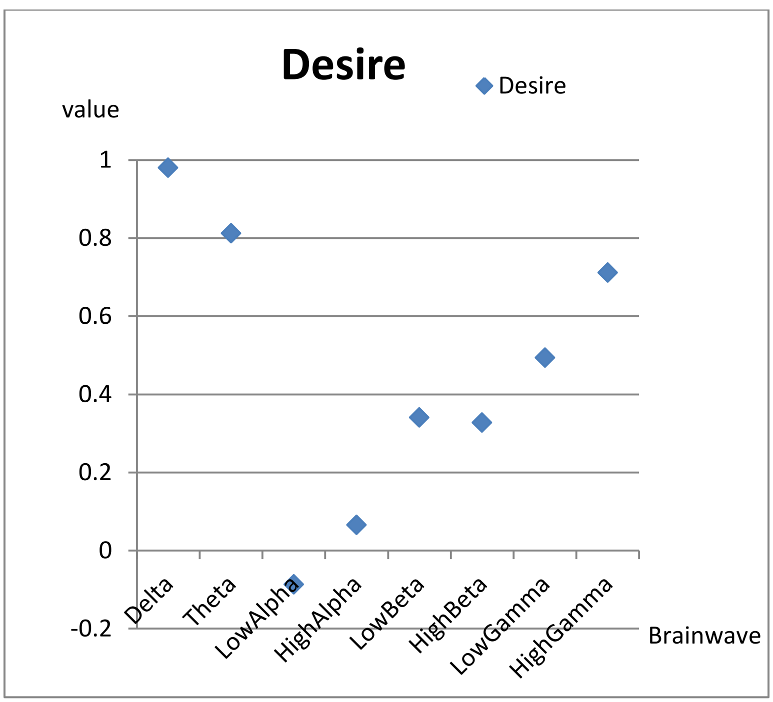


Table 5. Each Desire-Emotion Correlation

Desire / Emotion	Watch Movie	Eat the Cookie	Play the Game	Use Computer	See the Photo	Listen to the Music	Read Comics	Wild Card
Neutral	0.269	0.004	0.420	**0.655**	0.273	**0.870**	**0.939**	**0.825**
Happy	**0.567**	**0.530**	**0.535**	0.001	0.004	0.002	0.004	0.002
Surprised	0.115	0.40	0.024	0.287	**0.618**	0.062	0.042	0.134
Angry	0.012	0.028	0.001	0.001	0.002	0.004	0	0
Disgust	0.001	0	0	0.012	0.015	0.017	0.003	0.009
Fear	0.034	0	0.008	0.002	0.022	0.001	0.009	0.002
Sad	0.004	0.002	0.005	0.032	0.043	0.037	0	0.022

Figure 12 and Table 5 are shown the correlation between each desire and emotion.



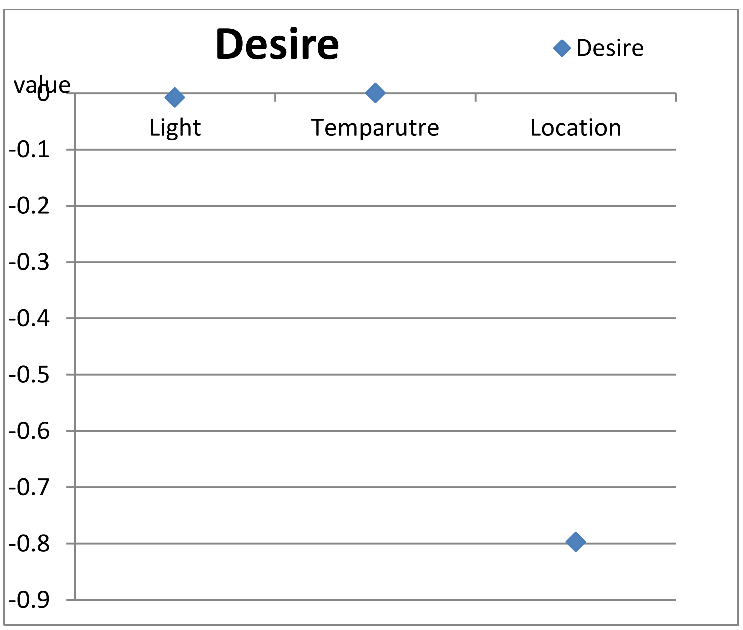
Correlation	Desire
Delta	***0.980***
Theta	***0.812***
LowAlpha	-0.087
HighAlpha	0.065
LowBeta	0.340
HighBeta	0.327
LowGamma	***0.493***
HighGamma	***0.711***

Figure 13. Correlation Result between Desire and Brainwave

Correlation	Desire
Behavior	***0.983***

Behavior

Figure 14. Correlation between Desire and Behavior



Correlation	Desire
Light	-0.008
Temperature	0
Location	***-0.797***

Env. Context

Figure 15. Correlation between Desire and Environmental Context

Based on our experimentation, we obtained several results as follows:

1. Desire inference can be derived from three types of parameters: Emotions, Behaviors, and Environments Contexts.
2. Affective States can be derived from two types of recognition schemes: Facial recognition and Brainwave.
3. Three basic emotions (Neutral, Happy, and Surprised) in the OCC model can be indicative to desire inference. It means that clear emotions can help to derive desire.

In the pilot study, we allowed exception of light and temperature sensors for the environment. Sometimes, these environmental contexts are very important factors depending on design.

B. Relationship between Emotion and Brainwave

The electroencephalograph (EEG) measures brainwave of different frequencies within the brain. Normally, there are five brainwave types including Delta (0-3Hz), Theta (4-7Hz), Alpha (8-12Hz), Beta (12-30Hz) and Gamma (30-50 Hz). Each brainwave type provides an emotional state and the corresponding conditions [12, 46, 47]. Table 6 shows the relationship between Emotion and Brainwave.

Table 6. Relationship between Emotion and Brainwave [12, 46]

Brainwave type	Frequency range	Mental states and conditions
Delta	0-3Hz	Deep, Dreamless sleep, non-REM sleep, Unconscious
Theta	4-7Hz	Joy (Happy), Fantasy, Creative, Intuitive, Normal (Neutral)
Alpha	8-12Hz	Normal (Neutral), Calm (tranquil), Relaxed but not drowsy
Beta	12-30Hz	Sadness, Anger, Disgust, Alertness, Agitation, Surprised, Thinking, Aware of self & surroundings
Gamma	30-50Hz	Fear, Disgust

Thus, an emotion recognition method using EEG signals is used [12]. EEG signals are measured and analyzed using power-spectrum analysis method based on the Fast Fourier Transform (FFT). Each EEG signal was decomposed into five brainwave types (EEG sub-bands) as mentioned before. It is straightforward to identify the different power rates at each frequency band as emotion. Such this discrepancy was investigated by analyzing human brainwaves. Correspondingly, each EEG signal is calculated by a relation power value equation that selects the EEG sub-band over the total EEG band:

$$\text{Relative power value } (\%) = \frac{\sum Selecting\ range}{\sum Total\ range\ (4\sim 50Hz)}$$

Then, the calculated EEG signals were in direct comparison with the EEG database. EEG signals were analyzed by Bayesian Networks (BNs). For Bayesian Network structure, we used Netica software [51].

Table 7 shows example Bayesian network structures. In comparing the results, many emotions showed different probability values, but “Anger” and “Sadness” show similar probability values.

Table 7. Relationship between Emotion and Brainwave [12]

– Results of calculation using Bayesian Network Structure for Emotion Recognition

Brainwave Type	Emotions	Brainwave Type	Emotions	Brainwave Type	Emotions
Low Theta	10% Joy 10% Fear 10% Anger 10% Disgust 60% Sadness	Low Theta	50% Fear 50% Disgust	High Theta	40% Normal 40% Sadness 4% Joy, Fear Anger, Disgust Surprised,
Low Gamma		High Gamma		Low Gamma	
Low Beta		Low Beta		High Beta	
Low Alpha		Low Alpha		Low Alpha	
Brainwave Type	Emotions	Brainwave Type	Emotions	Brainwave Type	Emotions
Low Theta	50% Anger 50% Disgust	Low Theta	50% Anger 50% Surprised	High Theta	70% Disgust 10% Fear 10% Anger 10% Surprised
Low Gamma		Low Gamma		High Gamma	
High Beta		High Beta		High Beta	
Low Alpha		Low Alpha		High Alpha	

Normally, Delta waves are easily collected from the brainwave in close association with the deepest stages of sleep, i.e., slow-wave sleep. We can also obtain Delta wave from unconsciousness and noise like pulse, neck movement and eye blinking. In reference [12], they explained, "We removed the delta band to eliminate EEG artifact (noise) such as pulses, neck movement, and eye blinking”. Along this line, Wakako Nakamura proposed the application of independent component analysis (ICA) together with the post processing of high-pass filtering to remove ballistocardiogram artifacts [47].

Thus, we also need to remove Delta from the results to get emotions because Delta type comes from unconsciousness. Also, the study of brainwave and certain belief showed a relationship as follows: *"The REM-dream state is a neurologically and physiologically active state. When a person is in deep sleep there is no dreaming and the waves (called delta waves) come at high amplitude about 3 per second. Dreams can be indirectly regarded as our unconscious feelings or/and desires very similarly with sleeping. We may have anxieties or desires that only our dreams can reveal. Most of us would have little difficulty in finding examples of "anxiety dreams" or "wish-fulfillment dreams" from our own experience. We may not have been aware of our desires or fears until we are awakened by the dream [48, 49]."*

Even though the delta band was removed, we wanted to show the whole results. From the high correlation results (Theta, Low Gamma, and High Gamma), we can infer Happy, Neutral, Fear and Disgust. Also, in our results, high beta and low beta (Surprised, Sadness, and Anger) have medium correlation.

Then, we calculated the average correlation between facial emotion and brainwave. As described in Table 5, we got results (Neutral, Happy and Surprised) of high correlation between desire and emotion factors. Brainwave is very different depending on individual conditions. We may need to have different groups to get more accurate results in the future.

CHAPTER 7. EVALUATION

In this chapter, we present the evaluation results on our approach. To validate our desire inference computational model, we have established an evaluation environment within the Smart Home Laboratory. During the experiments, multiple stimuli were introduced over time to induce changes in user's desires so that pilot data could be collected to train and verify the desire inference model.

Figure 7 in Chapter 5 schematically describes an overview of the sensor-rich and controlled environments in SHL at Iowa State University. In the experimentation room, multiple diverse sensors, software, and appliances as well as webcam/camcorder were deployed on difference objects. The dataset was collected for 3 weeks by 24 participants. Provisioning of stimuli was done without artificial control or causing intrusiveness that may influence the participant's activities during the experimentation. The desire state space (i.e., definition of desire) include desire for watching TV, playing a game, eating cookies, using a computer, seeing photos, reading comics, listening to music, and wild cards. The uses of wild cards help capture obscure or ambiguous desires.

7.1 Implementation

The computational model was implemented by MATLAB with Bayes Net Toolbox (BNT) introduced by Kevin Murphy [40]. BNT supports many types of conditional probability distribution, decision and utility nodes, static and dynamic BNs, and many different inference algorithms such as junction tree and frontier algorithm. For the inference algorithm, we used *the junction tree engine*. We first took the learning (training) data from 11 participants for constructing a desire inference model, and then selected the inference data from remaining 10 participants.

For the learning parameters, we used Expectation-Maximization (EM) algorithm as we mentioned in Chapter 4 [35]. EM algorithm is an iterative method to search maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models. This model depends on unobserved latent variables. The EM iteration consists of performing an expectation (E) step, which creates a function for calculating the expectation of the log-likelihood, evaluated using the current estimate for the parameters, and maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. Theses parameter-estimates are used to determine the distribution of the latent variables in the next E step [35]. Figure 16 shows E step, which uses an inference algorithm to compute the expected sufficient statistics, and M step in MATLAB.

```
engine{end+1} = jtree_dbn_inf_engine(bnet);
engine{end+1} = jtree_unrolled_dbn_inf_engine(bnet);

ncases = number of examples
max_iter= iterations for EM
```

E-Step

```
cases = cell(1, ncases);
for i=1:ncases
  ev = sample_dbn(bnet, T);
  cases{i} = cell(slice size,T);
  cases{i}(onodes,:) = ev(onodes, :);
end
[bnet2, LLtrace] = learn_params_dbn_em(engine, cases, 'max_iter', n);
```

M-Step

Figure 16. EM algorithm

EM algorithm can display the results after iterations of EM. Figure 17 shows the results after iterations of EM. Normally, complexity of EM algorithm depends on number of iterations and time to compute E and M steps.

The results after 2 Iterations of EM

```
EM iteration 1, loglik = -27.4714
EM iteration 2, loglik = -25.5910
engine 1 took 0.0760 seconds
EM iteration 1, loglik = -27.4714
EM iteration 2, loglik = -25.5910
engine 2 took 0.1338 seconds
EM iteration 1, loglik = -27.4714
EM iteration 2, loglik = -25.5910
engine 3 took 0.1727 seconds
EM iteration 1, loglik = -27.4714
EM iteration 2, loglik = -25.5910
engine 4 took 0.1713 seconds
iteration 1, loglik = -27.471430
iteration 2, loglik = -25.590991
```

Figure 17. EM algorithm iteration results

To evaluate our framework, there are several possible evaluation methodologies. First, we can have a simulation system in order to assure the correctness and effectiveness of the proposed model using the SHL facilities; thereby the feasibility of our framework and the effectiveness using synthetic and subjective data can be evaluated. Second, we can use a cross-validation route to the generalization of our training and test datasets. This technique can be beneficial for assessing how the results of a statistical analysis generalize an independent set of collected data. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model performs in practice [41]. We can evaluate the performance using the time-slice accuracy regarded as a typical technique in time-series analysis. It is noteworthy that the time-slice accuracy represents the percentage of correctly labelling time slices. Third, we can evaluate and compare the accuracies of different probabilistic models such as Hidden Markov Model (HMM) and Conditional Random Field (CRF). HMMs are generative models, not directly designed to maximize the performance of sequence labelling, thereby leading to modelling of the joint distribution with given observation. In contrast, CRFs based on a discriminative method are undirected graphical models, which are specially designed and trained to maximize performance of sequence labelling, thus resulting in modelling of the conditional distribution. Generally, discriminative methods are recognized to be more accurate since they are trained for a specific performance task. The principal difference of this approach with respect to the HMM is that it maximizes a conditional probability of labels given an observation sequence [31]. Fourth, we can use precision and recall measurements. In this dissertation, we selected this methodology for evaluation.

7.2 Precision and Recall

Precision and recall are the basic measures used in evaluating search strategies. Precision is defined as the fraction of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved, whereas recall is the fraction of the number of relevant records retrieved to the total number of relevant records. Both precision and recall are therefore originated from an understanding and measure of relevance [54].

Table 8 numerically demonstrates the results of the desire inference performance containing two parts of cells; the first numbers indicate adaptation of ours computational desire inference model with testing data and the second numbers correspond to the probability of Think-Aloud.

Comparison of these two numerical values allowed us to calculate the accuracy of our computational model.

Table 8. The Desire Inference Performance

Desire	**Recall (%)**	**Precision (%)**
Want to watch TV	97 88	96 98
Want to play a game	72 65	73 95
Want to eat cookies	63 72	95 93
Want to use a computer	81 92	87 89
Want to see photos	76 43	77 21
Want to read comics	81 23	77 33
Want to listen music	84 79	90 99
Wild Cards	71 84	81 90
Average	**78%** **73%**	**85%** **81%**

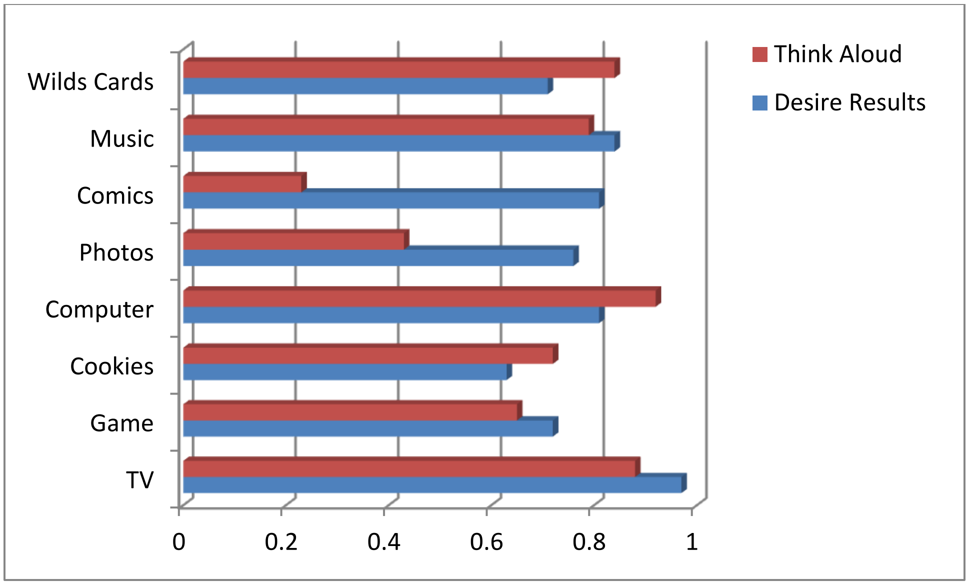


Figure 18. Recall of Desire Inference

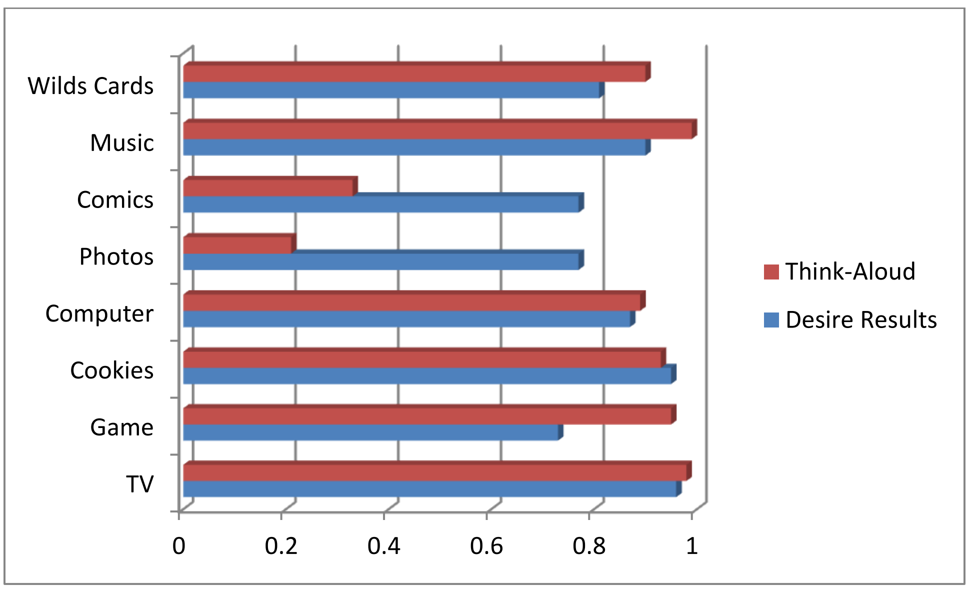


Figure 19. Precision of Desire Inference

In Figure 18 and Figure 19, the first (blue) bar shows the results after the application of our desire inference model with the testing data collected during the experiments. The second (red) bar of each desire shows the results of Think-Aloud.

Table 9. The Desire Inference Performance – Remove Emotions

Desire	Recall (%)	Precision (%)
Want to watch TV	92 88	97 98
Want to play a game	67 65	87 95
Want to eat cookies	60 72	96 93
Want to use a computer	78 92	92 89
Want to see photos	74 43	89 21
Want to read comics	78 23	88 33
Want to listen music	81 79	93 99
Wild Cards	72 84	86 90
Average	**75%** **71%**	**91%** **84%**

Desire	Neutral	Happy	Surprised	Angry	Disgust	Fear	Sad
80	0.999	0	0	0	0	0	0
80	0.999	0	0	0	0	0	0
80	0.999	0	0	0	0	0	0
80	0.999	0	0	0	0	0	0
80	0.999	0	0	0	0	0	0
80	0.999	0	0	0	0	0	0
20	0.999	0	0	0	0	0	0
20	0	1	0	0	0	0	0
20	0	1	0	0	0	0	0
20	0	1	0	0	0	0	0
20	0	1	0	0	0	0	0
20	0	1	0	0	0	0	0
20	0	1	0	0	0	0	0
20	0	1	0	0	0	0	0
20	0	1	0	0	0	0	0
20	0	1	0	0	0	0	0
20	0	1	0	0	0	0	0
20	0	1	0	0	0	0	0

Desire	Neutral	Happy	Surprised	Angry	Disgust	Fear	Sad
40	0	0	1	0	0	0	0
40	0	0	1	0	0	0	0
50	0.999	0	0	0	0	0	0
50	0.999	0	0	0	0	0	0
50	0.999	0	0	0	0	0	0
50	0.999	0	0	0	0	0	0

Figure 20. Examples of Desire Change Based on Emotion

Figure 20 continued

Desire	Neutral	Happy	Surprised	Angry	Disgust	Fear	Sad
80	0	0	0.98	0	0	0	0.02
80	0	0	0.999	0	0	0	0.001
80	0	0	1	0	0	0	0
80	0	0	1	0	0	0	0
80	0	0	1	0	0	0	0
80	0	0	1	0	0	0	0
40	0	0	1	0	0	0	0
40	0	0	1	0	0	0	0
40	0	0	1	0	0	0	0
40	0.999	0	0	0	0	0	0
40	0.999	0	0	0	0	0	0
40	0.999	0	0	0	0	0	0
40	0.999	0	0	0	0	0	0
40	0.999	0	0	0	0	0	0
40	0.999	0	0	0	0	0	0
40	0.999	0	0	0	0	0	0
40	0.999	0	0	0	0	0	0
40	0.999	0	0	0	0	0	0
40	0	0	0.988	0	0	0	0.012

Table 9 shows the desire inference performance after removing emotions. Although the average precision and recall of results are higher than the desire inference performance, we do not know about user's satisfaction. However, based on Figure 20, when emotion is changed, we know that the desire is changed, too. Thus, emotion is very significant factors whether the user is satisfied with their desire or not.

Desire inference results for wanting to watch comics and wanting to see photos deviated much from Think-Aloud results. The possible reasons for this inconsistency can be obtained from note-taking during the experiments as follows;

1) For the "wanting to watch comics" desire: When a participant wanted to watch comics using an iPad, it was not easy to distinguish what a participant really wanted versus actually announced through Think-Aloud. In other words, the participant's original desire was "watching comics," whereas the corresponding Think-Aloud answer could be "using an iPad."

2) For the "wanting to see photos" desire: Very similarly with case 1), some participants wanted to see photos using a computer. The participant's original desire was "wanting to see photos", whereas Think-Aloud answered "using a computer." As aforementioned, there would possibly be incoherency between real desire inference data and the corresponding Think-Aloud data.

CHAPTER 8. CONCLUSION AND FUTURE WORK

In this dissertation, we studied a desire inference process based on emotional, behavioral, and environmental context information. There are two major tasks as summarized below.

First, we presented a decision framework based on BBNs for modeling human desire using environmental, behavioral, and emotional contexts. Our current model takes considerations of three most frequently used modalities, including facial recognition, and brainwave analysis. In particular, the presented solution has offered two novel and critical contributions: (1) We gave a precise definition of the belief-perceived situation as B(m, a, e), and we demonstrated how to use belief-perceived situations to infer user's desire; (2) The probabilistic inference of human desire can be effectively carried out using a hierarchical DBBN.

Second, as fully described in the Chapter 3, we developed a theoretical desire inference model based on Bayesian belief networks (BBNs), which is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). Formally, Bayesian networks are directed acyclic graphs whose nodes represent random variables. They are observable quantities, unknown parameters or hypotheses. Efficient algorithms are applied for performing inference and learning in Bayesian networks. Bayesian networks that model sequences of variables are called dynamic Bayesian Belief Networks (DBBNs). Generalizations of Bayesian networks that can represent and solve decision problems under uncertainty are called influence diagrams [39]. The dynamic inference process based on DBBNs is being considered to allow the system to accommodate emerging new desires and adjust user's change in inclination towards certain desires over a period of time. DBBNs formalism is based on the Bayesian networks with extensions to represent discrete sequential systems.

Our central hypothesis is that DBBNs serve as an effective computational model for inferring user's desires when provided with real-time observations and finite historical data of the user's affective states, environmental contexts, and behavioural contexts as inputs. Based on our preliminary study, the problem of ambiguous human desire inference has been identified when human affective states were not considered. Therefore, the feasibility of using DBBNs to model human desires has been mathematically acquired. Thus, we completed an extended probabilistic desire inference model based on DBBNs and Shannon's information theory [3] and EM algorithm [35]. This two-layer hierarchical probabilistic framework is introduced to model the behavioural contexts, emotions, and environmental information largely related to the user's desire.

Third, we identified and verified key observable factors for effective and accurate user desire inference. We speculated that essential tasks for desire inference would be: (1) conducting the pilot study to collect human-centered dataset, and (2) establishing the correlation between the potential key factors and human desire.

It is essential to develop and evaluate a computational model capable of inferring users' desires using observable data captured either by the observing software or from the surrounding environments, such as those encountered in smart homes.

In addition, we established an evaluation environment within the SHL to validate our approach. During the experiments, multiple stimuli are disclosed over time to induce changes in user's desires so that pilot data are collected to train and verify the desire inference model.

Fourth, we presented the results of our framework using the data collected in this study. For the experiments with real (i.e., not simulated) data, we applied our framework to infer human desire. We used half of the data as our training data for constructing the computational model, and the other half as our testing data. The model implementation is in MATLAB using the Bayes Net Toolbox (BNT) [40]. BNT supports 1) many types of conditional probability distribution, 2) decision and utility nodes, 3) static and dynamic BNs and 4) many different inference algorithms such as junction tree and frontier algorithm.

We used the recall and precision methodology that is basic measures used in evaluation. Each desire has recall and precision percentage. Average precision was calculated to be 84% for human desire inference. Then, we showed the results the desire inference performance after removing emotions. The results were higher than the original desire inference performance. However, this comparison described that emotion is very important factor to change human's desire. The users changed their desire based on emotion's change.

In the future, we will investigate various approaches to improving the desire inference model.

First of all, other models such as CRF can be applied in the desire inference process. CRFs are undirected graphical models which belong to the discriminative class of models. CRF is based on a discriminative method, and specially designed and trained to maximize performance of sequence labeling. CRF models the conditional distribution. Usually discriminative methods are more accurate since they are trained for a specific performance task. This would possibly endow some better discernment to more complicated interactions between users and their surroundings.

Second, we can refine the experimentations. We have conducted the pilot study within the SHL. However, this was a first attempt. Based on the drawbacks of our pilot study, we can reinforce and extend the experimentation. We can consider different subjects with more varied background. Currently, the pilot study was very simple because of limitation of data-logging methodology. We need to come up with diverse methodologies to extend to different subjects. Also, we can have different setting with more sophisticated devices. Different software, sensors and devices can be considered.

Third, we may need to trim outlyers from the collected data, which was not done in our current experimentation. It is very important to generate large datasets in the field of HCI. There are several open source datasets for sharing. However, it is not satisfied with the average of the gold standard. Thus, the various approaches of the data collections are needed to propose.

In our pilot study, we had exception of constraints such as light and temperature sensors. Sometimes, they will be very significant factors to determine a result. All known constraints will be exploited in the future work.

REFERENCES

[1] M. Bauer, P. Gmytrasiewicz, and J. Vassileva, "User Modeling", User Modeling 2001: 8th International Conference (UM 2001), Vol. 2109, 2001.

[2] X. Li, "Integrating User Affective State Assessment in Enhancing HCI: Review and Proposition", Open Cybernetics and Systemic Journal, 2, pp. 192-205, 2008.

[3] X. Li and Q. Ji, "Active Affective State Detection and User Assistance with Dynamic Bayesian Networks", IEEE Transactions on Systems, Man and Cybernetics, Vol. 35, No. 1, pp.93-105, 2005.

[4] G. F. Schueler, "Desire: Its Role in Practical Reason and the Explanation of Action", Cambridge: MIT Press, 1995.

[5] A. Ortony, G. Clore, and A. Collins, "The Cognitive Structure of Emotions", Cambridge: Cambridge University Press, 1988.

[6] C. K. Chang, H. Jiang, H. Ming, and K. Oyama, "Situ: A Situation-Theoretic Approach to Context-Aware Service Evolution", IEEE Transactions on Services Computing, Vol. 2, No. 3, pp. 261 -275, 2009.

[7] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, and D. Riboni, " A survey of context modeling and reasoning techniques", Pervasive and Mobile Computing, 6, pp. 161-180, 2010.

[8] A. Masood Khattak, L. TheVinh, P. Tran Ho Truc, L. Xuan Hung, D. Guan, Z. Pervez, M. Han, S. Lee, and Y. Lee, "Context-aware Human Activity Recognition and Decision Making", 12th IEEE International Conference on e-Health Networking Applications and Services (Healthcom), pp.112-118, 2010.

[9] J. Muncaster and Y. Ma, "Activity Recognition using Dynamic Bayesian Networks with Automatic State Selection", IEEE Workshop on Motion and Video Computing (WMVC '07), pp.30-37, 2007.

[10] E. Nazerfard, B. Das, L. Holder, and D. J. Cook, "Conditional Random Fields for Activity Recognition in Smart Environments", Proceedings of the 1st ACM International Health Informatics Symposium (IHI '10), pp. 282-286, 2010.

[11] R. W. Picard: "Affective Computing", Cambridge, MA: The MIT Press, 1997.

[12] E. Ko, H. Yang, and K. Sim, "Emotion Recognition using EEG signals with Relative Power Values and Bayesian Network", International Journal of Control, Automation and System, Vol. 7, No. 5, pp. 865-870, 2009.

[13] I. Cohen, A. Garg, and T. S. Huang, "Emotion Recognition from Facial Expressions using Multilevel HMM", In Neural Information Processing Systems Workshop on Affective Computing, Vol. 2, 2000.

[14] 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 and User-Adapted Interaction, Vol. 12, pp.1-47, 2002.

[15] Y. Qi and R. Picard, " Context-Sensitive Bayesian Classifiers and Application to Mouse Pressure Pattern Classification", 16th IEEE International Conference on Pattern Recognition, Vol. 3, pp. 448-451, 2002.

[16] G. Ball and J. Breese, " Emotion and Personality in a Conversational Agent", Embodied Conversational Agents, pp. 189-219, 2000.

[17] C. Conati, "Probabilistic Assessment of User's Emotions in Educational Games", Applied Artificial Intelligence, Vol. 16, No.7-8, pp.555-575, 2002.

[18] A. Shashua, Y. Gdalyahu, and G. Hayun, "Pedestrian Detection for Driving Assistance Systems: Single-frame Classification and System Level Performance", Proceedings of IEEE Intelligent Vehicle Symposium, pp. 1-6, 2004.

[19] G. Fischer, "User modeling in Human-Computer Interaction", User Modeling and User-Adapted Interaction, Vol. 11, Issue 1-2, pp. 65-86, 2001.

[20] A. Kobsa, "Generic User Modeling Systems", User Model and User-Adapted Interaction, Vol. 11, Issue 1-2, pp. 49-63, 2001.

[21] P. E. Agre and S. J. Rosenschein, "Computational Theories of Interaction and Agency", Cambridge, MA: MIT Press, 1996.

[22] G. I. Webb, M. J. Pazzani, and D. Billsus, "Machine learning for user modeling", User Modeling and User-Adapted Interaction, Vol. 11, pp. 19-29, 2001.

[23] E. Hudlicka, "To feel or not to feel: The Role of Affect in Human–Computer Interaction", International Journal of Human-Computer Studies, Vol. 59, No. 1, pp.1–32, 2003.

[24] T. Bohnenberger, B. Brandherm, B. GroBmann-Hutter, D. Heckmann, and F. Wittig, "Empirically Grounded Decision-Theoretic Adaption to Situation-Dependent Resource Limitations", Künstliche Intelligenz, Vol. 16, No. 3, pp. 10-16, 2002.

[25] E. Horvitz, J. Breese, D. Heckrman, D. Hovel, and K. Rommelse, "The Lumiere Project: Bayesian User Modeling For Inferring the Goals and Needs of Software Users", Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp.256-265, 1998.

[26] J. Fieser and B. H. Dowden, “The Internet Encyclopedia of Philosophy”, Internet Encyclopedia of Philosophy Pub., 2001.

[27] R. W. Picard, E. Vyzas, and J. Healey, “Toward Machine Emotional Intelligence: Analysis of Affective Physiological State.”, IEEE Transactions on Patter Analysis and Machine Intelligence, Vol. 23, No.10, pp. 1175-1191, 2001.

[28] M. I. Jordan, “Learning in Graphical Models”, Cambridge, MA: The MIT Press, 1999.

[29] D. Heckerman, “A Tutorial on Learning With Bayesian Networks”, Technical Report, MSR-TR-95-06, Microsoft Research, 1995.

[30] Z. Lin and L. Fu, “Multi-user Preference Model and Service Provision in a Smart Home Environment”, Proceedings of the 3rd Annual IEEE Conference on Automation Science and Engineering, pp.759-764, 2007.

[31] Y. Chen, C. Lu, K. Hsu, L. Fu, Y. Yeh, and L. Kuo, “Preference Model Assisted Activity Recognition Learning in a Smart Home Environment”, The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.4657-4662, 2009.

[32] http://www.cs.cmu.edu/~dst/Tutorials/Info-Theory/

[33] V. Mihajlovic and M. petkovic, “Dynamic Bayesian Networks: A State of the Art”, 2001.

[34] J. Cheng, D. A. Bell, and W. Liu, “Learning Belief Networks from Data: An Information Theory Based Approach”, Proceedings of the Sixth ACM International Conference on Information and Knowledge Management, pp. 325-331, 1997.

[35] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from Incomplete Data via the EM Algorithm”, Journal of the Royal Statistical Society, Series B (Methodological), pp. 1-38, 1977.

[36] J. Dong, H. Yang, K. Oyama, and C. K. Chang, “Human Desire Inference Process Based on Affective Computing”, 34th Annual IEEE Computer Software and Applications Conference, pp.347-350, 2010.

[37] I. Hay, “Qualitative research methods in human geography (2nd ed.)”, Oxford: Oxford University Press, 2005.

[38] http://en.wikipedia.org/wiki/Correlation_coefficient

[39] K. Murphy, “Dynamic Bayesian Networks: Representation, Inference and Learning”, A Dissertation of Ph.D, University of California, Berkeley, 2002.

[40] https://code.google.com/p/bnt/

[41] L. Wang, T. Gu, X. Tao, H. Chen, and J. Lu, “Recognizing multi-user activities using wearable sensors in a smart home”, Pervasive and Mobile Computing, Vol. 7, No. 3, pp.287-298, 2011.

[42] C. Melo, P. Carnevale, and J. Gratch, “ The Influence of Emotions in Embodied Agents on Human Decision-Making”, 10th International Conference on Intelligent Virtual Agents (IVA 2010), LNAI 6356, pp. 357-370. 2010.

[43] N. Ponomareva, P. Rosso, F. Pla, and A. Molina, “Conditional Random Fields vs. Hidden Markov Models in a biomedical Named Entity Recognition task”, Proceedings of International Conference Recent Advances in Natural Language Processing, RANLP, pp.479-483, 2007.

[44] http://www.statisticshowto.com/articles/what-is-the-pearson-correlation-coefficient/

[45] P. Ekman and R. J. davidson, “The nature of emotion: Fundamental questions”, Oxford: Oxford University Press, 1994.

[46]http://developer.neurosky.com/docs/lib/exe/fetch.php?media=mindset_instruction_manual.pdf

[47] W. Nakamura, K. Anami, T. Mori, O. Saitoh, A. Cichocki, and S. Amari, “Removal of ballistocardiogram artifacts from simultaneously recorded EEG and fMRI data using independent component analysis,” IEEE Trans. Biomed. Eng., Vol. 53, No. 7, pp. 1294-1308, 2006.

[48] http://www.skepdic.com/dreams.html

[49] http://vajra.com.ar/english/bosu/bosutra6.htm

[50] http://www.r-project.org/

[51] https://www.norsys.com/index.html

[52] M. E. Bartman, “Intention, plans, and practical reason”, Harvard University Press, 1987

[53] P. Refaelizadeh, L. Tang, and H. Liu, ‘ Cross-Validation’, Encyclopedia of Database Systems, Springer US, pp. 532-538, 2009.

[54] http://www.creighton.edu/fileadmin/user/HSL/docs/ref/Searching_Recall_Precision.pdf

[55] J. J. C. Smart, “The identity theory of mind”, Stanford Encyclopedia of Philosophy, 2008.

[56] C. E. Shannon, “A Mathematical Theory of Communication”, The Bell System Technical Journal, Vol. 27, pp.379-423, 1948.

[57] T. Kirkland and W. A. Cunningham, “Mapping Emotions Through Time: How Affective Trajectories Inform the Language of Emotion”, Emotion®, Vol. 12, No. 2, pp. 268-282, 2012.

[58] K. M. Opdahl, “Emotion as Meaning: The Literary Case for How We Imaine”, Bucknell University Press, 2002.

[59] P. Vijay, MS. Bhatia, J. Sriniwas, and D. Batra, “Emotions and Mood”, Delhi Psychiatry Journal, Vol. 14, No. 2, pp.220-228, 2011.