Construction of a Dialog System with Emotions for Elderly Persons by Neural Networks

Takumi ICHIMURA
Hiroshima City University
Hiroshima, Japan

Kazuya MERA
Hiroshima City University
Hiroshima, Japan

Toshiyuki YAMASHITA
Tokyo Metropolitan Institute of Technology
Tokyo, Japan

Abstract

We developed an interactive system with dialog function that can analyze emotions of the user and response to him/her with facial expressions. The dialog system is due to our Emotion Generating Calculations (EGC) method based on the Elliot's Emotion Eliciting Condition Theory. The proposed method can judge whether an event is pleasant or not and obtain each attribute value for 20 various emotions under that situation. Successfully, we consider adjective relation and adverbial phrases and we input their attributes into the EGC, such that we can evaluate strength of each emotion.

This system can depict the face according to the context of conversation. Although each attribute value in EGC shows one aspect of feeling, we must integrate those attributes to express human-like emotions. We applied a fuzzy reasoning model for selecting the face proposed by T.Yamashita. In this method, nine faces with three levels of brow and eye deflection and three levels of mouth deflection were used to construct the fuzzy reasoning model. In this study, we extended Yamashita's system with eight emotions to a sophisticated system with 20 emotions by neural network.

1 Introduction

As computers have come into wide use, not only younger generation but also elderly one have a chance to use a computer. Elderly people, however, tend to think the use of the computer to be difficult. We asked the elderly subjects to state what kind of functions needs in a computer they try to use. As a result, they hope the computer equipped with human-like interfaces, which enables an easy conversation like the greeting and so on. Besides verbal messages, human face-to-face communication usually has nonverbal messages such as facial expressions, vocal inflection, speaking rate, pauses, hand gestures, body movements, posture, and so on. Therefore, one of such human interfaces must be the interface that can display the human face with some feeling under the conversation situation.

We developed an interactive system with dialog function that can analyze emotions of the user and response to him/her with facial expressions. The dialog system is due to our Emotion Generating Calculations (EGC) method[1][2] based on the Elliot's Emotion Eliciting Condition Theory[3]. The proposed method can judge whether an event is pleasant or not and obtain each attribute value for 20 various emotions under that situation. Successfully, we consider adjective relation and adverbial phrases and we input their attributes into the EGC, such that we can evaluate strength of each emotion.

This system can depict the face according to the context of conversation. Although each attribute value in EGC shows one aspect of feeling, we must integrate those attributes to express human-like emotions. We applied a fuzzy reasoning model for selecting the face proposed by T.Yamashita[4]. In this method, nine faces with three levels of brow and eye deflection and three levels of mouth deflection were used to construct the fuzzy reasoning model. In this study, we extended Yamashita's system with eight emotions to a sophisticated system with 20 emotions.

If an elderly person is introversive during conversation, he/she often feels depressed. Even if such case is occurred, the proposed dialog system can analyze the condition in his/her emotions and change the topics and get his/her feeling so good. The content of response is also determined to select one sentence from prepared database. However, when the selected face or sentence is different from the response of real occurring emotion, they are adaptively corrected by neural network. The neural network learns the difference between estimated emotion and user's emotion in real conversation content.

To verify the effectiveness of the proposed method, we tried to use the method for the analytical system of health service needs of elderly in an elderly subjects' home[6]. We asked him/her to evaluate the usability of our system and we discussed the results for further development of our system.

0-7803-6583-6/00/$10.00 © 2000 IEEE

2 Emotion Generation Calculations

The study of spoken natural language dialog has been of practical interest and quite a few dialog systems have successfully been developed within the specified task domains. However, human-computer interactions of currently designed systems are not satisfactory enough. Incorporating emotions into the dialog processing of such systems should lead to more human-like performance. In this section, we present a method for constructing an emotion-handling dialog system, in order to facilitate more effective interaction with the users. We describe how to generate emotions from the utterances, focusing on the similarities between the grammar structures and the meaning structures within the case frame. We can reduce the number of emotion-generating expressions by sharing those of the same meaning structures. Word impressions are used for their variables. Furthermore, we apply these expressions to the negatives and the noun phrases. The result of questionnaire expressions showed that the relation of word-impressions and emotions is almost good.

2.1 Agent Models

We depict an agent model in Figure 1 to communicate human and agents. The agent model consists of mainly three parts;
(1)Analyses of input utterance(sentence analyses domain),
(2)Decision of what-to-say (dialog planning domain),
(3)Decision of how-to-say(sentence generation domain).
In addition, the agent model has internal world for short term memory such as self-model, other's model and current situation, memory management domain for long term memory such as knowledge and episode, and a emotion generation domain for various events.

When the other utterance sentence is given from the external world, the sentence analyses domain in the system analyzes the content of the utterance and the utterance intention. As for dialog planning domain in the system, the content of the answer is decided based on the intention of other utterance, conversation career, and present feelings.

Moreover, the domain judges the situation and decides the answer method. Based on the decided content of the answer, the sentence generation domain in the system generates character strings related to a concrete utterance and outputs to the external world in consideration of the other party model. The emotion-generating domain exerts the influence on all these process.

2.2 Emotion Generating Calculations

We consider that human emotion is consisted of two parts; "emotion" and "mood". The "emotion" means feelings rapidly caused by a certain event and "mood" expresses continued feelings, which change by various "emotions." We can usually categorize human feelings into "pleasure," "sadness," "angry," "expectation", and so on. Such these feelings are based on whether the event is pleasant/unpleasant for some people. The proposed EGC method gives the measure of a degree of his/her pleasant/dispersant/unknown to the event when an input is given in the form of the case frame representation into emotion generating domain in the system. In this paper, we define the equation of Emotion Generating Calculation for a type of case frame respectively, as in the table 1. The 12 case frame structures in the table 1 are summarized as follows.
Type I) Subject(*S*) does Verb(*V*) that influences reach to subject's own.
Type II & III)*S*'s statement that has relation to *V's* changes from Object-From(*OF*) to Object-To(*OT*) by *S*.
Type IV) *S* and *OM* have relation to *V*.
Type V) *S* and Object-Source(*OS*) do *V* at the same time.
Type VI) *S* does *V* to Object(*O*).
Type VII & VIII) *O* is done *V* by *S*.
Type X) Instrument(*I*) is also multiplied by the type VI-b) of EGC because it has relation to the emotion strength.
Type XI) *O* has an attribute Object-Content(*OC*).
Type IX & XII)*V*s in these type *S* are various qualities. So this type of EGC cannot be summarized into one category.

Table1 Event type and Emotion Generating Calculation

Type	Event type	Emotion Generating Calculation
I	$V(S)$	$f_S \times f_P$
II	$V(S,OF)$	$f_S \times (f_{OT} - f_{OF}) \times f_P$
III	$V(S,OT)$	$f_S \times (f_{OT} - f_{OF}) \times f_P$
IV	$V(S,OM)$	$f_S \times f_{OM} \times f_P$
V	$V(S,OS)$	$(f_S - f_{OS}) \times f_P$
VI	$V(S,O)$	a) $f_S \times (f_O \times f_P)$ b) $f_O \times f_P$
VII	$V(S,O,OF)$	$f_O \times (f_{OT} - f_{OF}) \times f_P$
VIII	$V(S,O,OT)$	$f_O \times (f_{OT} - f_{OF}) \times f_P$
IX	$V(S,O,OM)$	———
X	$V(S,O,I)$	$f_O \times f_P \times f_I$
XI	$V(S,O,OC)$	$f_O \times f_{OC}$
XII	Others	———

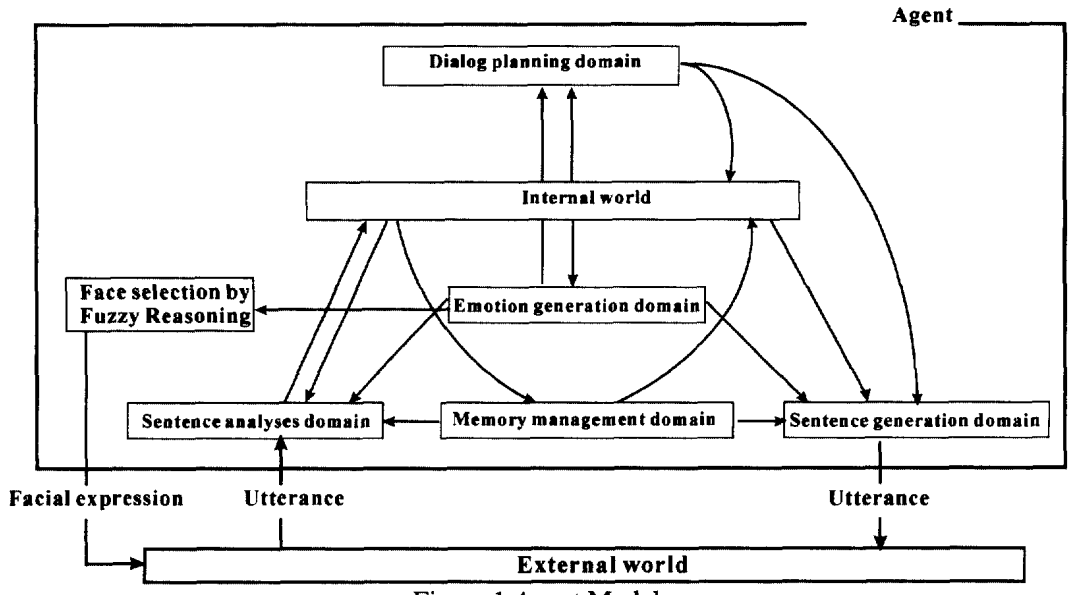

Figure 1 Agent Model

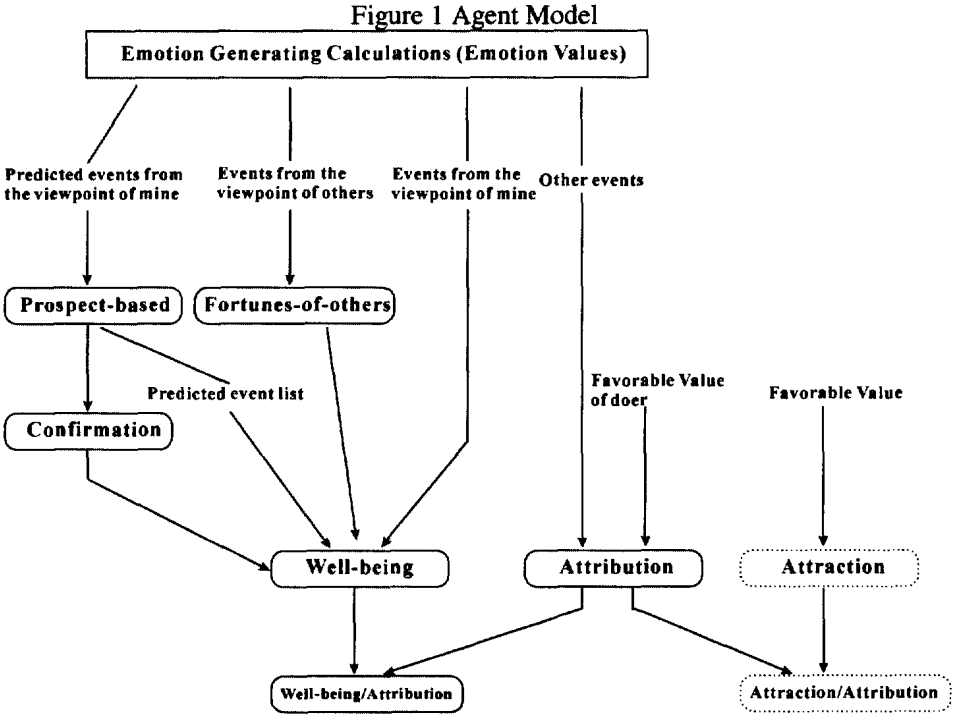

Figure 2. Dependency of Emotion Generation

In the table 1, the f s are defined in the following.

f_S Favorable Value for "Subject"

f_O Favorable Value for "Object"

f_{OF} Favorable Value for "Object-From"

f_{OT} Favorable Value for "Object-To"

f_{OM} Favorable Value for "Object-Mutual"

f_{OS} Favorable Value for "Object-Source"

f_{OC} Favorable Value for "Object-Content"

f_P Favorable Value for "Predicate"

An Emotion value by EGCs is given from the inner product value of each favorable value obtained by dividing one sentence. To obtain the degree of favorable value, we measure the strength of the emotions from elements except favorable value. Therefore, we consider the following point of view to evaluate the degree of strength of emotions; "Some elements related to only strength are added to EGCs", "Pick up the attribute except favorable values", and "Calculate corresponding degree to the modification structure." We assume an attribute value of the remarkable element for each event and assign the appropriate attribute value for the object.

Based on such emotion values and their situations, the degree of pleasure/displeasure for each emotion type is calculated. In this paper, we consider only 20 emotion types. Elliott defines 24 emotion types in [3], that is, "joy," "distress" as a group of "Well-Being"; "happy-for," "gloating," "resentment," "sorry-for" as a group of "Fortunes-of-Others"; "hope," "fear" as a group of "Prospect-based"; "satisfaction," "relief," "fears-confirmed," "disappointment" as a group of "Confirmation"; "pride," "admiration," "shame," "reproach" as a group of "Attribution"; "liking," "disliking" as a group of "Attraction"; "gratitude," "anger," "gratification," "remorse" as a group of "Well-being/Attribution"; "love," "hate" as a group of "Attraction/Attribution." The emotion type "liking" and "disliking" are not included in the generated emotion type by EGC. Since the emotion type "love" and "hate" are different from others, we will discuss in [7].

Figure 2 shows the dependency between the groups of emotion type.

Let us present an example calculation.

Event : "Ryota dates with Kiriko."

$$\begin{bmatrix} Predicate(P) & = & \text{"}dates\ with\text{"} & :+0.5 \\ Subject(S) & = & \text{"}Ryota\text{"} & :+1.0 \\ Object-Mutual(OM) & = & \text{"}Kiriko\text{"} & :+0.1 \end{bmatrix}$$

Event Type: "date with" $\rightarrow V(S,OM)$

$\downarrow$

Emotion Value

$= f_S \times f_{OM} \times f_P$

$= f_S(I) \times f_{OM}(Kiriko) \times f_P(dateswith)$

$=(+1.0)\times(+0.1)\times(+0.5)$

$=+0.05 \rightarrow$ positive number(a little pleasant)

The result shows that the agent feels a little pleasant about the event "Ryota dates with Kiriko."

3 Construction of the model

3.1 Faces and Emotional Affections

We used nine facial expressions with three levels of brow deflection and three levels of mouth deflection, as shown in Figure 3.

3.2 Fuzzy Reasoning Model

We assumed that the ratio of the subjects who selected the face for a given affection was the grade to which the face belonged to a fuzzy set of faces which expressed the affection.

As shown in Figure 4, we proposed the following fuzzy reasoning model on the basis of Yamashita[4][5]:

Rule 1 : $A_1 \Rightarrow P_1$

Rule 2 : $A_2 \Rightarrow P_2$

. . . .

Rule20 : $A_{20} \Rightarrow P_{20}$

Fact : a_1 and a_2 and $\cdots$ a_{20}

Conclusion : P' ,

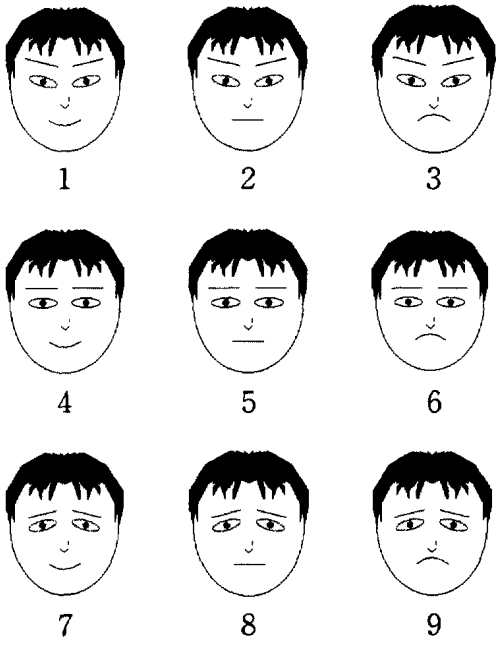

Figure 3. Nine facial expressions

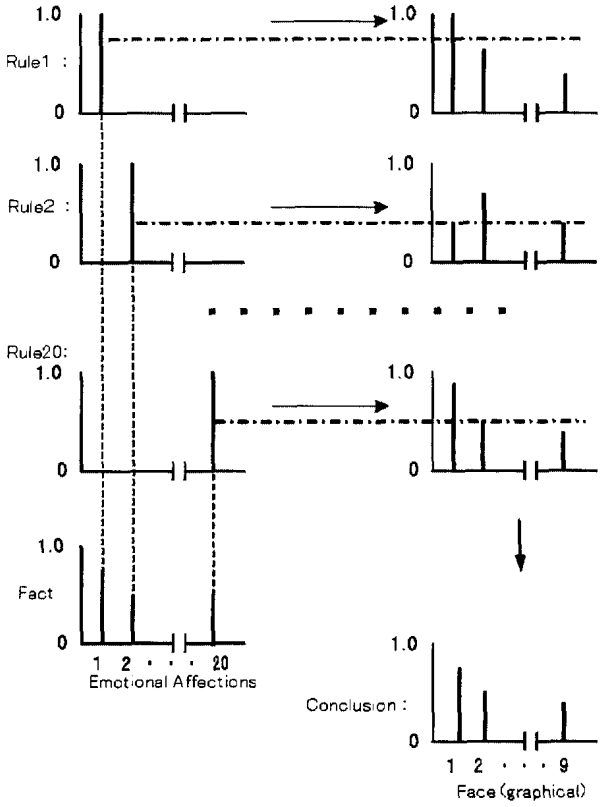

Figure 4. Fuzzy reasoning model

where A_1, A_2, ···, and A_{20} in the antecedent part are fuzzy sets of emotional affections, and P_1, P_2, ···, and P_{20} in the consequent part are fuzzy sets of faces which express a given emotional affection.

If we have a_1, a_2,···, and a_{20} as the intensity of each emotional affection, then the fuzzy reasoning result of Rule i (i=1, 2,···,20) is calculated by

$$\mu_{Pi'}(z) = a_i \wedge \mu_{Pi}(z),$$

and the combined conclusion of Rule 1, Rule 2, ···, and Rule 20 is given as

$$\mu_{P'}(z) = \mu_{P1'}(z) \vee \mu_{P2'}(z) \vee \cdots \vee \mu_{P20'}(z).$$

The faces with the highest or the second highest grades of membership should be selected as the face which well expresses the emotional affections caused by a given situation.

4 Decision Model by Neural Networks

In this paper, we developed a dialog system using neural network. We have adopted neural network learning for the following reason. As for cause-consequence relationship between the cognitive state and the physiological one, neural network is considered as the appropriate framework for expressing the relationship based on the sample datasets. Although the estimated values differ from person to person, neural network can express the suitable performance for the expression of pleasant/dispersant for each person.

The configuration of the neural network for emotion recognition and decision face type is shown in Figure 4. This network is a combination of general neural network and the decision logic stage. The network evaluates output activities of nine output neurons and determines the final recognition result. Each output neuron corresponds to a face type and the reply sentence accompanied with the face is determined. The pair of the face type and the reply sentence is a predetermined rule in dataset.

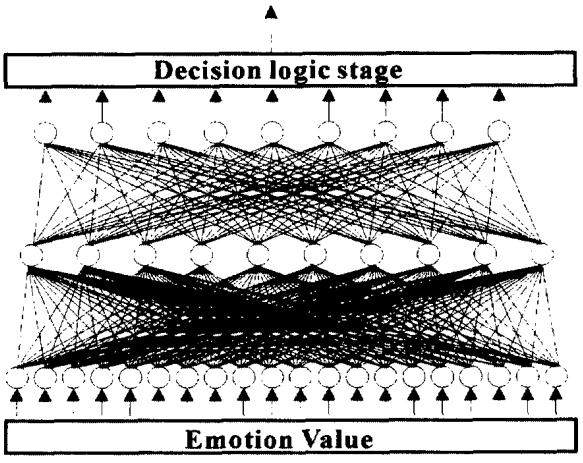

Figure 4. Decision facial type model by neural networks

The system takes advantages in the following point. If an elderly person is introversive during conversation, he/she often feels depressed. The content of response is also determined to select one sentence from prepared database. However, since the selected face or sentence is different from the response of real occurring emotion, they are adaptively corrected by neural network. The neural network works to learn the difference between estimated emotion and user's emotion in real conversation content.

5 Experimental Results

5.1 WWW-based Analytical System of Health Service needs among Healthy Elderly[6]

An increase of elderly persons requires not only medical cases but also health services. The Japanese society has already built up the medical insurance and medical care systems. However, the health service needs in so-called healthy elderly are not analyzed. We developed the www-based analytical system of health service needs of elderly. The basic questionnaire consisted of 50 items. These items related to QOL, -Quality of Life-, life-satisfaction, life-style, mental stress, and social concern. The subject tried to answer these questionnaire sheets in the homepage and these answers are sent to website. Successfully, the system checks if the questionnaire were fulfilled all conditions and were stored to implement a reasoning. Based on such calculations, the system presented the analytical results and health counseling comment to the subject in the browser.

The system was also developed to analyze the population-based health service needs to the official health center. The health service for the elderly persons will be offered on the basis of these results. The system is expected to classify the health service needs of the elderly in his/her home.

Although the personal computer diffusion in Japan reaches up to about 30% according to the statistical data of Economic Planning Agency of Japan, not a few people have difficulties with the use of computer. Especially, we often see that the elderly tends to think the use of the computer to be difficult. This kind of problem is caused by the distance between the conversation and the usability of the computer. That is, they hope the computer equipped with human-like interfaces, which enable an easy conversation like the greeting and so on. Besides verbal messages, human face-to-face communication usually nonverbal messages such as facial expressions, vocal inflection, speaking rate, pauses, hand gestures, body movements, postures, and so on. In

this study, to improve this weak point of the developed www-based health service system, we devised the method of analyzing subject's answer.

5.2 Training Data for Dialog System

For example, there are two pairs of question and answer on the questionnaire sheets; Question 1 and Question 2. The appropriate EGC from their sentences and the emotion type are determined and the emotion value is calculated by the EGC simultaneously.

Question1: Do you lose your way occasionally?
Answer: I might lose one's way if going out of the city.

Emotion Value = $f_S \times f_P = 0.8 * (-0.5) = -0.4$

Emotion Type={"sorry-for", "fear", "distress"}

Face Type=8

Based on such results by EGC, we obtained the following training data.

{0.0, -0.4, 0.0, 0.0, 0.0, -0.4, 0.0, -0.4, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8}

Question2: Do you think that the forgetfulness became cruel recently?
Answer: I do not do the one forgetting at all.

Emotion Value =

$f_S \times (f_O \times f_P) = 0.8 * (-0.7) * (0.5 * 1.5 * (-1.0)) = 0.42$

Emotion Type={"joy", "happy-for", "admiration"}

Face Type=1

Based on such results by EGC, we obtained the following training data.

{0.42, 0.0, 0.42, 0.0, 0.0, -0.4, 0.0, -0.4, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.42, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1}

The training data set is obtained for each person and the network is trained to represent his/her feeling. In Figure 4, if the decision logic stage recognizes a certain face type which he/she feels depressed, the response in this system changes the question. The reply sentence is also determined to select one sentence from the prepared database.

6 Conclusion

We developed an interactive system with dialog function that can analyze emotions of the user and response to him/her with facial expressions. The dialog system is due to EGCs. However, EGC cannot analyze a subject's affirmation/negative intention from an answer. We have already developed the method of analysis of user communicative intention from affirmation/negative elements to improve such weak point[8].

Moreover, we should use the predetermined rule database related to the system's reply for a predicted subject's answer when the subject feels depress. Although, the database is desired for each subject, it is not easy to investigate many person emotions, since the situation may change according to various unpredicted occasion. The neural network with traditional BP learning cannot ensure a good performance under such a situation. We consider that the adaptive learning method by neural networks takes advantage under the variant situation[9]. We will apply the method to the dialog system in the near future.

References

[1]K.Mera, "An emotion calculate method by word-feeling fro natural language dialog", Technical report of IEICE, NLC98-26, pp.1-8(1998)(In Japanese)

[2]K.Mera, "Revision of Emotion Invoking Calculations in order to provide strength of emotions, Technical report of IEICE, TL99-32, pp.47-54(1999)(In Japanese)

[3]C. D. Elliott, "The Affective Reasoner: A process model of emotions in a multi-agent system", A Doctor Dissertation, Northwestern Univ.(1992)

[4]T.Yamashita, M.Takahashi, H.Sakai, T.Taketa and T.Ichimura, "An Applications of Facial Selection Model by Fuzzy Reasoning to Human Interface", Journal of Japan Society for Fuzzy Theory and Systems, Vol.12, No.2, pp.313-320(2000)

[5]T. Yamashita, "Fuzzy Reasoning model of facial selection and its applications", Proc. of the 5th Intl. Conf. on Soft Computing and Information/Intelligent Systems, 201-204(1998)

[6]K. Yoshida, T. Ichimura et al., "Analytical System of Health Service needs among Healthy Elderly by using Internet", Proc. of Geontechnology Third Intl. Conf.(1999)

[7]K.Mera, S.Kawamoto, Mitsuko Yamamura-Takei and Teruyuki Aizawa, "Emotion-based planning evaluation method", Accepted for Fourth Intl. Conf. on Knowledge-Based Intelligent Engineering Systems & Allied Technologies(2000)

[8]K.Mera, T.Ichimura and T.Yamashita, "Analysis of User Communicative Intention from Affirmation/ Negative Elements by Fuzzy Reasoning and Its Application to WWW-based Health Service System for Elderly", submitted to the 6th Intl. Conf. on Soft Computing(IIZUKA2000)(2000)

[9]S.Oeda, T.Ichimura, M.Terauchi, T.Takahama and Y.Isomichi, "Adaptive Evolutional Learning Method of Neural Networks using Genetic Algorithms under Dynamic Environments", Accepted for Fourth Intl. Conf. on Knowledge-Based Intelligent Engineering Systems & Allied Technologies(2000)