Soft Computing-based Emotion/Intention Reading for Service Robot

Z. Zenn Bien[1], Jung-Bae Kim[2], Dae-Jin Kim[2], Jeong-Su Han[2] and Jun-Hyeong Do[2]

Div. of EE, Dept. of EECS, KAIST 373-1 Guseong-Dong, Yuseong-Gu,
Daejeon 305-701, KOREA
[1] zbien@ee.kaist.ac.kr,
[2] {jbkim, djkim, pyodori, jhdo}@ctrsys.kaist.ac.kr

Abstract. Due to its tolerance of imprecision, uncertainty and partial truth, the soft computing technique deals well with human related signals such as voice, gesture, facial expression, bio-signal, etc. In this paper, we propose architecture of soft computing-based recognition for a class of biosign. Especially, the problem of inferring emotion and intention reading from such recognition is considered with applications for service robots that interact with human. It is shown that proposed architecture renders good performance in a few experimental systems for rehabilitation.

1 Introduction

The service robots are mainly designed to serve humans directly or indirectly by helping or replacing humans in the works that usually require human flexibility under unstructured, possibly varying environments and sometimes intense-interactions. They may immensely differ from the industrial robots that repeat only those works predefined in a structured workspace.

The service robots take various forms and functions. For examples, they include housekeeping home robots, entertainment robots, rehabilitation robots for the disabled, intelligent robot house, etc. For these service robots, an important basic technology which needs a special attention is "human friendly interface" including voice recognition, gesture recognition, object recognition, user's intention reading, etc. This technique focuses on human-machine interaction because the service robots receive direct human command or cooperate with human.

To recognize *biosigns* such as voice, gesture, facial expression and bio-signals, we need an intelligent recognition method that is tolerant of imprecision, uncertainty and partial truth of biosign. Here, bio-signals include ECG (Electrocardiogram: heart signal), EMG (Electromyogram: muscle signal), EEG (Electroencephalogram: brain signal), etc. The soft computing method, which differs from the conventional hard computing paradigm, is known to have those characteristics and potential to solve many real-world problems [8]. The soft computing techniques contain fuzzy logic, neural network, probabilistic reasoning, evolutionary algorithms, chaos theory, belief networks, learning theory, etc [9].

This paper proposes an architecture of a soft computing-based recognition system that deals with human mind level signals, especially the user's emotion and intention. Reading human emotion and/or intention is a challenging research topic, because there are apparently no appropriate sensors/measurement devices for many biosign-related variables, and, even if possible, subjectiveness of the result may have an ill effect.

The rest of the paper is organized as follows: Section 2 introduces emotion and intention and how to perceive them. Section 3 introduces more details about signal flow in man-machine interaction systems. In Section 4, we suggest an architecture of soft computing-based recognition system. We deal with application systems in Section 5. This paper concludes in Section 6.

2 Preliminary

2.1 Emotion/Intention

The word 'emotion' is used very often in our daily lives. According to [5], it is very difficult to answer the question such as 'What is the emotion?' because of its wide usage and subjective characterization [6]. However, we use the term 'emotion' to express our natural feeling of happiness, joy, sadness, surprise, anger, greeting, love, hate and so on. In this paper, the word 'emotion' is also used to represent such feelings as well as mood and affect [7].

Intention is an act or instance of determining mentally some action or result. It is a direct representation of the user's purpose, whereas emotion is an indirect one. For example, "bringing the cup to the user's mouth" is a good example of direct representation of the user's purpose, and we may relate it with an intention of the user. On the other hand, a negative reaction such as "shutting the user's mouth when the robot serves" may be interpreted as an emotional state to express that the user does not want to eat anything, which may be interpreted as a kind of indirect representation of the user's purpose, and we may relate it with emotion of the user.

From a psychological point of view, there have been many attempts to understand "how a human can recognize emotions/intentions of the other humans" Mehrabian proposes an emotion-space model called "PAD Emotional State Model" [14]. It consists of three nearly independent dimensions that are used to describe and measure emotional states: *Pleasure-displeasure, Arousal-nonarousal, and Dominance-submissiveness.* "Pleasure-displeasure" distinguishes the positive-negative affective quality of emotional states, while "arousal-nonarousal" refers to a combination of physical activity and mental alertness. And "dominance-submissiveness" is defined in terms of control versus lack of control. Visual stimuli-based approach by Ekman et al. is also very popular. They proposed that many emotions or intentions in human's face may be recognized by combination of various facial muscular actions so called "AU (Action Unit)" [15]. Dellaert et al. attempted to find elements that can affect emotions from speech signals [16].

On the basis of these psychological approaches, many researchers have been also trying to recognize human emotions (or intentions) for engineering purpose. An emotional agent proposed by Breazeal can recognize human's emotions based on PAD emotional state model [17]. This agent can recognize and represent many emotions based on PAD emotional model with mechanical structures. Vision-based approaches based on Ekman's theory shows promising results [4]. With soft computing techniques, machine can effectively recognize human's emotions based on acquired facial expression images. And Nicholson gave a shot to recognize emotions from speech signals using artificial neural networks [18].

2.2 Soft Computing Tool Box

Soft computing techniques are convenient tools to solve many real world problems. It is known to exploit the tolerance for uncertainty and imprecision to achieve tractability, robustness, and low solution cost [9]. Key methodologies include the fuzzy logic theory (FL), neural networks (NN), evolutionary computation (EC), and the rough set theory (RS), etc. Complementary combination of these methodologies may show a strong computing power that "parallels the remarkable ability of the human mind to reason and learn in an environment of uncertainty and imprecision [8]."

Two concepts play a key role within FL [9]. One is a *linguistic variable* and the other is a *fuzzy if-then rules*. FL mimics the remarkable ability of the human mind to summarize data and focus on decision-relevant information.

NN is a massively parallel-distributed processor made up of simple processing units, called *neurons*, which is a natural propensity for storing experiential knowledge and making it available for use. Nonlinearity of neuron, input-output mapping, adaptivity, and fault tolerance are useful properties of NN [19].

EC can be described as a two-step iterative process, consisting of *random variation* followed by *selection*. In the real world, EC offers considerable advantages such as adaptability to changing situations, generation of good enough solutions quickly, and so on [20].

By applying RS into a data set that is incomplete, imprecise, and vague, we can extract knowledge in a form of a minimal set of rules [21]. RS provides many advantages including efficient algorithms for finding hidden patterns in data, data reduction, methods for evaluating significance of data, etc.

To summarize, FL, NN, EC and RS can be appropriate tools for rule induction leaning, optimization and rule reduction, respectively.

3 Signal Flow in Man-machine Interaction System

Fig. 1 shows a model which we propose to describe signal flow from human's *mind level* to machine's *action decision making module*. Emotion and intention in mind level induce various biosigns through the many human's physical organizations such as face, hand, muscle, brain and vocal card in the *body level*. These biosigns include bio-signals, gesture, facial expression, voice, eye gaze, etc.

The machine senses biosigns using various sensors in *acquisition module* and recognizes emotion and (or) intention in the *emotion/intention reading module*. Finally, the machine's actions are made between human and service robots.

To deal with the biosign, which has imprecision, uncertainty and partial truth, soft computing tool box is used in emotion/intention reading module and *action decision making module*. The detailed part from the acquisition module to emotion/intention module is dealt in Section 4. As the man shows some biosign to the machine and the machine recognizes the biosign and gives action to the man, they make the man-machine interaction.

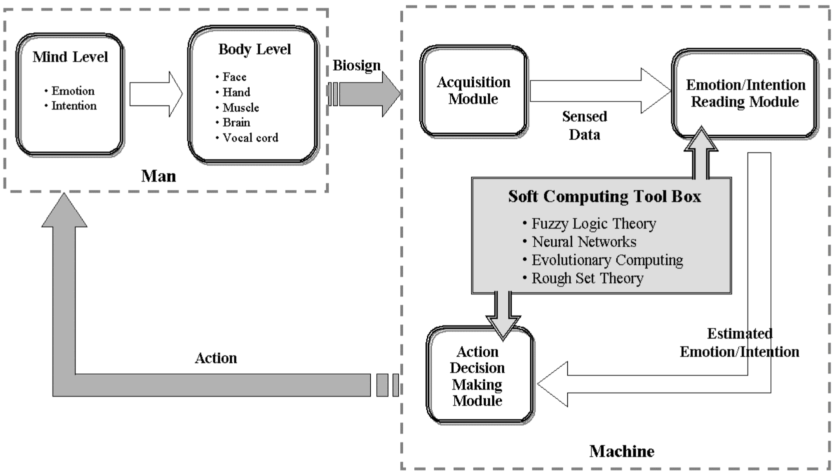

Fig. 1. Soft computing-based emotion/intention reading procedure from human mind level to action decision making level

4 An Architecture of Soft Computing-based Recognition System

As in cases of human, the partner's intention or emotion can be inferred not only from language but also from behavior. Typically, inferred intentions or emotions are vague and not necessarily expressible, but it plays a key role for conservative decision making as in the case of design in consideration of safety or for smooth cooperation for comfort. A human being also tries to read the other party's intention or emotion subjectively. Thus, any classical probability or statistics may not be appropriate to express one's intention or emotion in a mathematical way [1]. Hence, we need appropriate methods, such as soft computing techniques, to deal with these types of vague and uncertain knowledge.

We propose a soft computing-based recognition system for the biosign as shown in Fig. 2. It is a modified figure of the fundamental step of digital image processing [10]. The input of the architecture is biosign and the output is the recognized intention, emotion, information and exogenous event.

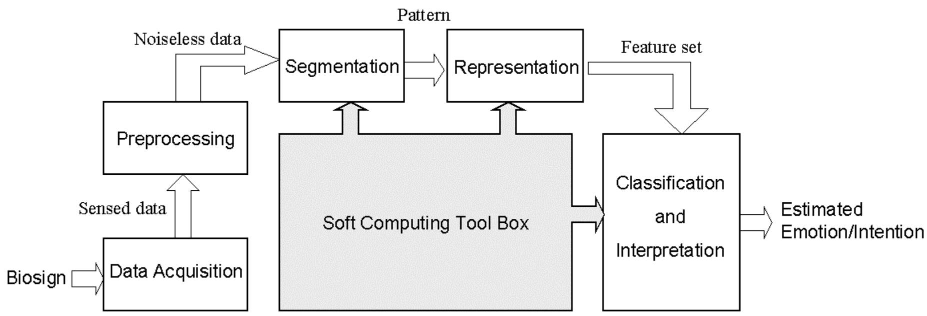

Fig. 2. An Architecture of soft computing-based recognition system

The starting block of the system is “data acquisition”-that is, acquiring biosigns. The sensors for acquisition could be microphone, camera, glove device, motion capture device, EMG signal detector, etc. After the biosign is obtained, the next step deals with preprocessing. The preprocessing block typically deals with enhancing the signal and removing noise. The next stage deals with segmentation. It means partitioning a biosign into constituent signals. In general, it contains two segmentation parts: spatial segmentation and temporal segmentation. The former means selecting meaningful signal from a signal mixed with background signal, and the latter means selecting isolated signal from a continuous signal.

The output of the segmentation stage needs to be converted into a form suitable for computer processing. This involves representation of raw data. It contains the feature extraction processing. The last stage of Fig. 2 involves classification and interpretation. Classification is the process that assigns a label to an object based on the information provided by its features. Interpretation involves assigning meaning to an ensemble of classified objects.

To deal with biosign, we need prior knowledge in the processing modules in Fig. 2. We implement it with soft computing technique. As we mentioned, FL, RS, EC and NN may be appropriate method for rule induction, rule reduction, optimization and learning respectively. So, we apply FL and NN to the segmentation stage, FL and RS to the representation stage, and FL, NN and EC to the classification and interpretation stage. As auxiliary methods, state automata and Hidden Markov Model are used for segmentation and classification stage.

5 Application Systems

5.1 Gesture Recognition System [11][12]

To overcome inconveniences of human-machine communication tools such as keyboards and mouse, the hand gesture method has been developed to accommodate a variety of commands naturally and directly. In spite of its usefulness, however, hand gesture is difficult to recognize by a machine.

Construction of a hand gesture recognition system involves structural categorization of gesture, real-time dynamic processing, pattern classification in a hyper dimensional space, coping with deterioration on recognition rate in case of expansion of gesture, dealing with ambiguity and nonlinearity constraints of the sensors, etc. Naturally, several intelligent processing methods such as soft computing technique have been evolved to overcome these difficulties. In our works, we use state automata to segment a continuous gesture into a set of individual gestures. And we use fuzzy min max neural network in the hand posture and orientation classification [11]. Also, we use FL and Hidden Markov Model in the hand motion classification [12].

5.2 Facial Emotional Expression Recognition System [4]

In general, the problem of recognizing emotion from a face is known to be very complex and difficult because individuality may come in expressing and observing emotions. It is interesting to note, however, that human beings can successfully understand facial expressions in a seemingly easy way. In our work, various soft computing techniques are used effectively for recognizing a positive expression of happiness [5]. This work has adopted NN, FS, and RS theory. To handle the recognition system by employing a traditional FL framework, a novel concept termed as "fuzzy observer" is proposed to indirectly estimate a linguistic variable from conventionally measured data.

As an application, we have been developing an intention reading system for intelligent visual servoing for wheelchair-mounted rehabilitation robot system, KARES II [2]. Among many principal tasks of KARES II, we focus on the task of "bringing the cup to the user's mouth". During this task, we use the mouth region to infer the user's intention such as a will to drink or not. According to the extracted user's intention, the robot may approach to the user's face or get away from it. A fuzzy rule base is constructed using linguistic variables such as 'mouth openness', 'positive intention', and 'negative intention'. Here, the fuzzy observer is also effectively used for extracting user's intention during service [3].

5.3 Bio-signal Recognition System [13]

The EMG control is well known from the operation of some prosthesis with small DOF. Its application to the user's high level of movement paralysis is limited because the useful signals often interfere with the EMG signals from another muscle groups. The soft computing technique allows effective extraction of informative signal features in cases of high interference between the useful EMG signals and another muscle EMG signals.

To read the user's movement intentions effectively, we have proposed the minimal feature set extraction algorithm [14] based on the fuzzy c-means algorithm (FCM), and RS. We can obtain the intervals of each feature by FCM to make condition rules, and then apply the rough set theory to extract a minimally sufficient set of rules for classification. After extracting numerous rules for classification and reduction done by RS, we can find the best feature set by measuring the separability of each feature in each rules. By use of fuzzy min-max neural network (FMMNN) as a pattern recognizer with the extracted minimal feature sets, we can classify the eight primitive arm motions with high success classification rates [14].

5.4 Service Robot System with Emotion Monitoring Capability [7]

To help human mentally and emotionally, a service robot system is designed to understand the user's emotion and react depending on the monitored information.

In our works, we build an intelligent robot agent for emotion-invoking action and emotion monitoring to combine the user's emotion and emotional model of the agent. For emotion monitoring, the robot is to observe the user's behavior pattern that may be caused by some change of surroundings or some initial robot action and to understand the user's emotional condition and then act by ingratiating itself with the user. For learning, the robot agent gets a feedback from the user's response so that it can behave properly, depending on any situation for the user's sake. Most important problem is to establish a mapping concept from the user's behavior pattern to the user's emotional state and from the emotional state to the robot's action ingratiating the user. Since each mapping rule depends on the personality of the user, it is difficult to determine universal affective properties in the user's behavior pattern and robot's action. By the proposed NN structure, it is described how the robot would understand the user's emotional condition and how it shows its reaction, depending on the user's emotional state as a service robot [7].

6 Concluding Remark

Soft computing techniques can deal with many real-world problems effectively. Among many possible applications of soft computing techniques, human-machine interface or interaction procedures for the service robots are found to be very suitable because of its capability to deal with uncertainty and ambiguity. In this paper, we have proposed a novel scheme for emotion/intention reading based on various soft computing techniques. And four successful applications are given as examples based on the proposed scheme.

Acknowledgement

This work was supported by the Human-friendly Welfare Robot System Engineering Research Center, KAIST.

References

1. Y. Inagaki, et al., “Behavior-based intention inference for intelligent robots cooperating with human,” Proc. of Int’l Conf. of the 4th FUZZ-IEEE, vol. 3, pp. 1695–1700, 1995
2. Z. Bien, and W.-K. Song, “Novel wheelchair-based robotic arm with visual servoing capability for human-robot interaction”, Proc. of Workshop on Service Automation and Robotics, CIDAM, pp. 5-17, June 19-20, 2000.
3. D.-J. Kim, W.-K. Song, and Z. Bien, "Effective Intention Reading in Rehabilitation Robitcs", In the Proceeding of HWRS2001 Workshop, KAIST, Korea, Jan. 15-16, 2001.
4. Gyu-tae Park and Zeungnam Bien, "Neural network-based fuzzy observer with application to facial analysis" , Pattern Recognition Letters, vol. 21, pp. 93-105, February, 2000.
5. Yong-Min Kim, “Emotion”(in Korean), Han-Gil co., 1999.
6. Eun-Kyung Cho et al., “Cognitive Psychology”, Hak-Ji co., 1999.
7. J.-H. Do, K.-H. Park, et al., “A Development of Emotional Interactive Robot,” Proc. of 32nd ISR2001, pp.544-549, Seoul, Korea, 2001.
8. Jang. J. S., et al., Neuro-fuzzy and soft computing, Prentice-Hall, 1997.
9. Zadeh L.A., Soft computing and fuzzy logic, IEEE software, vol. 11, pp. 48–56, 1994.
10. Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, Addison-Wesley Publishing Co. 1993.
11. J.-S. Kim, W. Jang and Z. Bien, “A Dynamic Gesture Recognition System for the Korean Sign Language (KSL)”, IEEE Trans. on Systems, Man, and Cybernetics, Vol. 26, No. 2, pp. 354-359, 1996.
12. J.-B. Kim, K.-H. Park, W.-C. Bang, J.-S. Kim and Z. Bien, “Continuous Korean Sign Language Recognition using Automata based Gesture Segmentation and Hidden Markov Model”, ICCAS2001 submitted, 2001.
13. J.-S. Han, Z. Bien, et al., “New EMG Pattern Recognition based on Soft Computing Techniques and Its Application to Control of a Rehabilitation Robotic Arm,” Proc. of 6th IIZUKA2000, pp.890-897, Iizuka, Japan, Oct. 1-4, 2000.
14. A. Mehrabian, “Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies,” Oelgeschlager, Gunn & Hain, Cambridge, MA, 1980.
15. P. Ekman and W.V. Friesen, The Facial Action Coding System, Consulting Psychologist Press, Inc., San Francisco, CA, 1978.
16. F. Dellaert, T. Polzin and A. Waibel, “Recognition emotion in speech,” Spoken Language, 1996.
17. C. Breazeal, “Robot in Society: Friend or Appliance?,” Agents’99 workshop on emotion-based agent architectures, pp. 18-26, 1999.
18. J. Nicholson, K. Takahashi and R. Nakatsu, “Emotion recognition in speech using neural networks,” Proceedings of ICONIP’99, vol. 2, pp. 495-501, 1999.
19. S. Haykin, Neural Networks, Prentice-Hall, 2nd edition, 1999.
20. D. B. Fogel, “What is Evolutionary Computation?”, IEEE Spec. pp. 26-32, Feb., 2000.
21. Z. Pawlak, “Why Rough Sets?”, Proc. Of the fifth IEEE Inter. conf. on. Fuzzy Systems, vol. 2, pp. 738-743, 1996.