Using Fuzzy Classifiers for Perceptual State Recognition

Ian Beaver *
School of
Computing and Engineering Science
Eastern Washington University
Cheney, WA 99004-2493 USA
ibeaver@mail.ewu.edu

Atsushi Inoue †
School of
Computing and Engineering Science
Eastern Washington University
Cheney, WA 99004-2493 USA
atsushi.inoue@ewu.edu

Abstract

We are studying perceptual state recognition such that the states in a state machine are identified in the "space of meaning", and recognized directly from only a few sensory information of variable quality. In this paper, we present the effect of using fuzzy classifiers for this type of recognition. We are mainly concerned with the feasibility of developing intelligent systems in a simple way, by showing two cases: image processing for a video conference system and generation of emotional behavior for a computer game.

Keywords: Computational Perception, Fuzzy Classifier, State Recognition.

1 Perceptual State Machines

In principle, computer systems perform their given tasks as a series of transitions among states, that are determined based on their input information, and, in many cases, outputs are generated in association with transitions or states (aka Mealy machines vs. Moore machines). Indeed, intelligent systems follow the same principle as well, except for various types of non-determinism that exists as their intrinsic characteristic. Such non-determinism is somehow well handled by human beings in their daily lives, yet mimicking task performance of human beings by computers is still a great challenge.

Indeed, many scientists and engineers are studying extensively in order to overcome this challenge. While a fundamental dilemma has always been confronted in many approaches necessary increase of system complexity and intrinsic complexity of intelligent tasks, Soft Computing approaches have demonstrated high potentials to break this dilemma[10, 11]. A notable approach of Soft Computing, among many others, is the consideration of perception as a basis of intelligent systems. Listing efforts of such an approach, Zadeh advocates and shows feasibility of *Computational Theory of Perception*[12], Ralescu and Shannahan studied a *perceptual object recognition from image* based on Gestalt Theory[1], and Inoue and Ralescu studied a method of *Perceptual Information Processing* and its application to text classification problems[6].

In those approaches, the following common framework exists: intelligent tasks are identified and presented in the "space of meaning" rather than the space of patterns on sensory information, and perception is represented by mapping from concepts in the space of meaning to sensory information and vice versa. Terano[8] suggested that use of natural language is the most appropriate (and natural) to identify and label those concepts, and

*Corresponding author. A graduate (master) student of Computer Science at EWU.

†The Chief Scientist for Intelligent Systems, EWU Center for Network Computing And Cyber Security (CNCACS). This work is in part supported by EWU's Technology Initiative for New Economy (TINE) grant.

this leads us to the issue of identifying appropriate mapping between linguistic labels and sensory information[8]. Further, Ralescu discussed this framework as a paradigm of image understanding[2].

From aspects of state machines –the principle of computer systems performing tasks– and within the framework of perceptual approaches, we consider the following *perceptual state machines* as a representation of intelligent systems:

1. **All states**(S) are identified in the space of meaning. They are labeled in a natural language (i.e. linguistic label).

2. **Transitions** (T) are established among those states.

3. **Inputs** (I) are generated through a compilation of sensory information.

4. **Outputs** (O) are optional, but can be generated according to Mealy machine and Moore machine.

When no outputs are generated (i.e. recognizer), we call this machine *Perceptual State Recognizer* and refer its execution (i.e. Computation) as *Perceptual State Recognition*. Likewise, we call this *Perceptual State Transducer* when generating outputs and refer its execution as *Perceptual State Transduction*.

More formally, transitions ($t \in T$ where T is a set of transitions) can be represented as functions such that

$$t : S \times I \rightarrow S$$

where S is a set of linguistic labels identifying states in the space of meaning, and I is a set of inputs through compilation of sensory information. Notice that this function is an instance of the mapping between linguistic labels and sensory information mentioned earlier.

In general, the utility of Perceptual State Machines is high as we can easily project properties of general state machines. The efficiency of intelligent systems represented in those machines is significantly dependent on the implementation and its performance of transition functions as well as the structural complexity (i.e. the dimensionality when considering a vector space) of input space, i.e. sensory information.

2 Fuzzy Classifiers as Transition Functions

One of the commonly known properties of fuzzy sets is the association of a linguistic label l and data domain D (numerical in many cases of sensory information) such that their membership functions are defined as

$$\mu_l : D \rightarrow [0, 1]$$

Consequently, fuzzy classifiers can be considered as transitions of Perceptual State Machines. Formally, a fuzzy classifier F_l as a transition function at state $l \in S$ can be represented as

$$F_l(x \in I) = arg\ max_{l \in S}[\mu_l(x)]$$

where membership functions μ_l are associated with transitions from state l to all other states, including oneself.

It is interesting to note that a perceptual state machine is deterministic when there is no tie on the membership values for an input x.

3 Video Conference System

As an example of perceptual state machines, we consider a video conference system[7] that autonomously controls audio and visual devices so that images and sound of students in remote classrooms who need the instructor's attention are appropriately transmitted to the main classroom where the instructor resides. We attempt to recognize three states presented in Table1 solely by analyzing images taken from streaming video of classroom activities (e.g. Figure1).

Membership functions of a fuzzy classifier are shown in Figure2. Such states (in a space of meaning) are recognized as a result of classifying signals of human movement (in a space

Table 1: Perceptual States

State	Description
IN_SESSION	A class is in session.
NO_SESSION	A class is not in session.
QUESTION	A student is raising a hand (the instructor's attention).

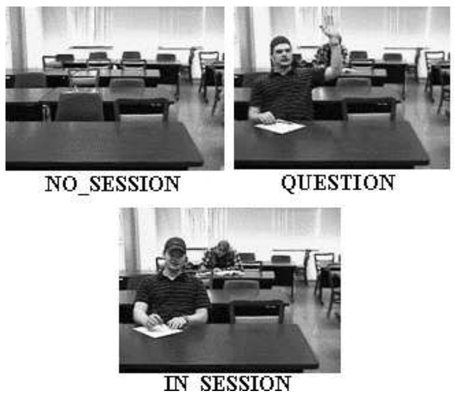

Figure 1: Sample Images

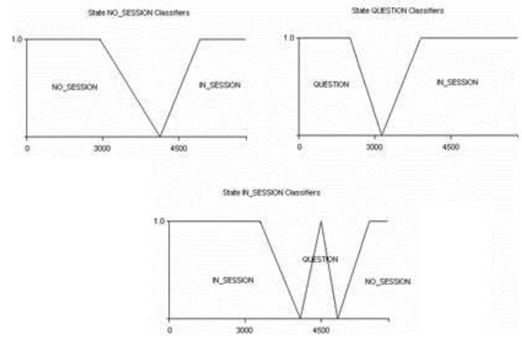

Figure 2: Fuzzy Classifiers

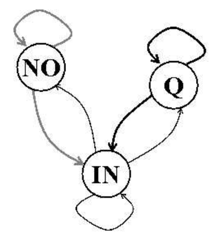

Figure 3: State Diagram

of patterns). Figure3 shows the state diagram. In this diagram, no transitions between 'NO_SESSION' and 'QUESTION' exist.

In our study, the signal of movement is captured by looking into the change of pixels. In particular, this is extracted by differentiating two images captured through a CCD camera within a short interval, normally between one to two seconds (i.e. feature extraction for classification).

It is important to note that the interval is crucial to the recognition. If the interval is too long, too much information will be missed, and this will not classify the state successfully. On the other hand, too little differentiation can be captured (and thus this classification is difficult) if the time interval is too short.

3.1 Feature Extraction: Inputs

We use a simple feature (i.e. a single dimensional) for representing the signal of movement. That is the number of pixels within a certain range of grayscale whose value is different between two images in the last two captures. Since we are interested in capturing movement of students, this range can be determined based on the color range of human body. Among parts of the human body, it is observed that the color range of human palms falls in a more specific range regardless of the pigment color of the rest of the body. Because movement of palms significantly contributes to recognize the state in which a student asks a question, we consider the color range of palms as the core feature.

While advances of technology make the CCD camera capture more information, e.g. a higher resolution and full color images in RGB format, all information may not be necessary for this recognition. Further, the less information this recognition task takes into account, the simpler its computation becomes. Consequently, we decided to use the grayscale between 0 and 255.

A conversion from RGB to grayscale can be performed based on the following simple linear combination:

$$\Xi_\psi = 0.299 \cdot \Xi_R + 0.587 \cdot \Xi_G + 0.114 \cdot \Xi_B$$

where Ξ_ψ is luminance; and Ξ_R, Ξ_G, and Ξ_B are the primary color signals of red, green and blue, respectively. Such a conversion has been implemented as a specific function in the Python Imaging Library[3].

Formally, the feature value $x \in I$ is obtained as follows:

$$x = \Sigma_{i,j} r(M_1(i,j)) - \Sigma_{i,j} r(M_2(i,j))$$

where i and j represent the coordination of a pixel (i.e. the i-th raw and the j-th column in a matrix M consisting of the luminance Ξ_ψ, representing a raster image in the grayscale). Function r filters pixel e such that

$$r(e) = \begin{cases} 0 & \text{if } e \text{ is out of range} \\ 1 & \text{otherwise} \end{cases}$$

3.2 Optimization of Fuzzy Classifiers

For further improvement of the performance, we look into a method of optimizing this classifier suitable for online problem domains such as this recognition problem. This optimization needs to take many factors that may cause significant influences to the performance into consideration. Those include, but are not limited to, lighting, clothing, and skin color. Considering the sensitivity of the CCD camera to lighting as an example, the optimization may be necessary for various situations as follows:

1. Strong sunshine comes in at a certain time period.
2. Lights are adjusted in the classroom.
3. Fixtures do not provide enough lighting.

The optimization of fuzzy classifier is achieved by altering membership functions (e.g. Figure2) as deemed necessary. In our work[7], we use Mass Assignment Theory (MAT)[4], a correspondence between a fuzzy set and a family of probability distributions, to reconstruct fuzzy sets as histograms of input values (i.e. $x \in I$) are updated.

3.3 Performance

To measure the performance, we have captured a video stream of a classroom setting with two students sitting: one at the front-most and the other at the back (as shown in Figure1). We have then manually labeled all transitions between any two frames in the video stream and used those labels in order to measure the accuracy. For each interval between two frames (i.e. 13 different intervals from 0.3 seconds to 4 seconds), we have evaluated the accuracy for five times (i.e. the first iteration through the fifth iteration). Between two iterations, we have optimized fuzzy classifiers based on MAT.

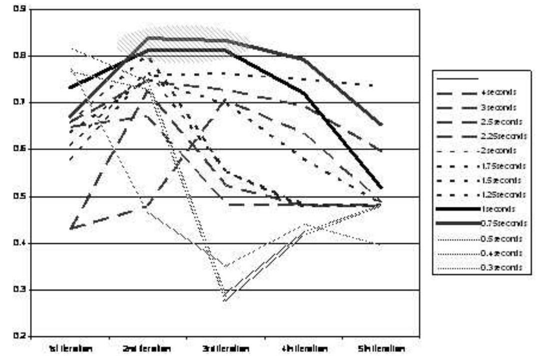

Figure 4: Accuracy of Recognition

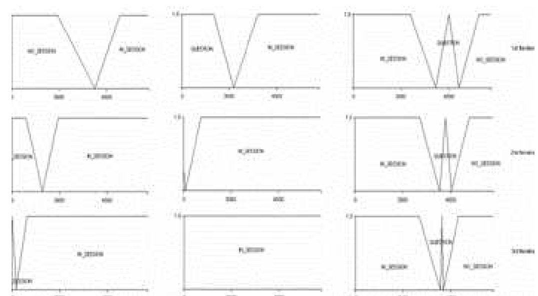

Figure 5: Optimization of Fuzzy Sets

Figure4 shows the accuracy of classification across the five iterations. As mentioned at the beginning, there is an optimal interval, i.e. around 1 second. The peak accuracy is around 84%. Notice also the optimality concerning the iteration. The second and the third iterations with the interval of 1 second (as highlighted in Figure4) appear to be the peak of performance. This means that one or two optimization of fuzzy classifiers have resulted in the peak, and many more eventually decrease the performance. In Figure5, the change of fuzzy sets as a result of optimization (based on MAT) is shown (fuzzy classifiers at 'NO_SESSION', 'QUESTION', and 'IN_SESSION' from the left to the right; the first through the third iterations from the top to the bottom).

Consequently, we have learned the following effects of using fuzzy classifiers as transition functions of this perceptual state recognizer:

- The structure in terms of the number of states is simple. So should be the development as a consequence.
- The accuracy in the first iteration (i.e. the case of perceptually configured fuzzy classifiers) is acceptable in many cases.
- Optimization improves the accuracy, but not significantly for cases with reasonably high accuracy, i.e. around 70% accuracy or higher in this study. In fact, the accuracy is likely decreased as if the optimization is performed in those cases.
- Use of such a simple sensor results in the accuracy of 84%. It is hopeful for a significant improvement with more sensory information as well as better sensors (i.e. a future work).

4 Generation of Emotional Behavior

Another example of perceptual state machine is to generate emotional behavior for a first-person shooter computer game, *Quake II*[5] shown in Figure6. Recently, there is a high demand on more attractive humanistic features within computer games, rather than the low-level artificial intelligence (i.e. path-finding details) such as believable and interesting non-player characters (NPCs) that perform complex reasoning and learning in order to exhibit emotions. We developed a perceptual state machine that generates such behavior on monster soldiers in the game. The soldiers then respond with humanistic (i.e. emotional) behavior depending on various inputs.

4.1 Perceptual State Transducer

This transducer consists of six emotional states (shown in Table2). For each state, there is an associated action of a monster soldier hard-coded within the source implementation as briefly mentioned in Table3. Transitions are determined based on the (fuzzy set) conjunction of a membership function of 'Aggregation' and that of 'Fear'. The domain of those membership functions are the Cartesian product of three variables (i.e. sensory inputs) as follows:

Figure 6: Quake II

- Health of NPC
- Distance between a NPC and the first-person shooter (i.e. player)
- Angle between a NPC and the shooter

Table 2: Emotional States

State	Aggregation	Fear
Psycho	high	low
Snipe	low	low
Cautious Advance	high	medium
Evasive Attack	low	medium
Panic	high	high
Afraid	low	high

Table 3: Emotional States

State	Action
Psycho	Sprint towards player and attack
Snipe	Shoot from a distance.
Cautious	Shoot and slowly move advance towards player.
Evasive	Jump from side-to-side and shoot.
Panic	Freeze and run in random directions (The NPC cannot do anything because of fear).
Afraid	Run in the opposite direction of the player.

4.2 Performance

Despite the simplicity, this perceptual state transducer adds a significant difference to the behavior of NPCs, i.e. monster soldiers, in comparison to the original Quake II (see the video clips in the presentation to verify this).

Further, we also found that a four-state perceptual transducer (consisting only of 'Psycho', 'Snipe', 'Panic' and 'Afraid') performs similarly well.

We also found that, through a trial of a simple reinforcement learning mechanism, adaptability of fuzzy classifier (such as the one based on MAT) adds nothing but unnecessarily complexity to the game playing (i.e. too difficult to play). This positively supports the effects of fuzzy classifiers for simple structure and development.

5 Concluding Remarks

The concept of Perceptual State Machine is introduced, and its effects, i.e. the *simplicity* in the structure and development, are presented through two case studies: the video conference system and the emotional behavior of NPCs in a computer game.

We are currently working on the following:

1. Development tools – API and methodologies of embedded system development
2. Continuation on the video conference system development for distance learning
3. Further study on sensors – fusion of multiple sensors, use of inexpensive sensors, etc.

References

[1] A. L. Ralescu, J. G. Shanahan, "Perceptual Organization for Inferring Object Boundaries", *Pattern Recognition*, 32.11, pp. 1923-1933, 1999.

[2] A. L. Ralescu, "Image Understanding = Verbal Description of the Image Contents", *Journal of the Japanese Society for Fuzzy Theory and Systems*, Vol. 7, No. 4, in Japanese, August 1995.

[3] F. Lundh, M. Ellis, *Python Imaging Library Overview - PIL 1.1.3*, `http://www.pythonware.com/products/pil/pil-handbook.pdf`, March 12, 2002.

[4] J.F. Baldwin, T.P. Martin, B.W. Pilsworth, *FRIL-Fuzzy and Evidential Reasoning in Artificial Intelligence*, Research Studies Press, 1995.

[5] T. Hooley, B. Hunking, M. Henry, A. Inoue, "Generation of Emotional Behavior for Non-Player Characters: Development of EmoBot for Quake II", *Proceedings of AAAI*, San Jose, CA, 2004.

[6] A. Inoue, A. L. Ralescu, "Computational Model of Perceptual Information Processing", *Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE99)*, Seoul, South Korea, pp. 824-829, 1999.

[7] I. Beaver, A. Inoue, "Perceptual Recognition of States in Remote Classrooms", *Proceedings of International Conference of North America Fuzzy Information Processing Society (NAFIPS05)*, Ann Arbor, MI, 2005.

[8] T. Terano, S. Masui, T. Terada, H. Watanabe, "Coloring of a Landscape by Fuzzy Logic", *Journal of the Japanese Society for Fuzzy Theory and Systems*, Vol. 5, No. 2, pp. 375-385, in Japanese, April 1993.

[9] A. Bargiela, W. Pedrycz, *Granular Computing*, Kluwer Academic, 2003.

[10] L. A. Zadeh, "Fuzzy Logic, Neural Networks, and Soft Computing", *Communication of the ACM*, vol. 37, No. 3, 1994.

[11] T. Terano, et. al. (ed), *LIFE Research Report*, Software AG, in Japanese, 1995.

[12] L. A. Zadeh, "From Computing with Numbers to Computing with Words: From Manipulation of Measurements to Manipulation of Perceptions", *IEEE Transaction on Circuits and Systems I*, vol. 45, No. 1, January 1999.