

# Mutual Information Analysis with Similarity Measure

Hongmei Wang and Sanghyuk Lee\*

School of Mechatronics, Changwon National University  
\*Institute for Information and Electronics Research, Inha University

## Abstract

Discussion and analysis about relative mutual information has been carried out through fuzzy entropy and similarity measure. Fuzzy relative mutual information measure (FRIM) plays an important part as a measure of information shared between two fuzzy pattern vectors. This FRIM is analyzed and explained through similarity measure between two fuzzy sets. Furthermore, comparison between two measures is also carried out.

**Key Words** : certainty; fuzzy relative mutual information; fuzzy entropy; shared information; similarity measure

## 1. Introduction

Major issue in unsupervised pattern recognition is the designing similarity between two vectors. The determining measures are referred to the similarity measure (SM) and dissimilarity measure (DM). These measures play an essential role in pattern recognition, classification and clustering. Information quantization represents interesting research theme, in which data vagueness can be illustrated by clear number. Design of fuzzy entropy for calculation of uncertainty has been studied by numerous researchers [1-3]. Most of results were concentrated in the designing of fuzzy entropies [1,2], and many parts of them also showed the implicit results of fuzzy entropies [1]. Hence, to apply real data explicit fuzzy entropy has to be needed. For information evaluation similarity measure has to be needed. Applying similarity measure, there must be needed comparing data sets. Conventional similarity measure has been designed based on the fuzzy number and distance measure [4-6]. Similarity measure with fuzzy number can be found in references [4]. However, similarity measure is restricted within the triangular and trapezoidal fuzzy membership function cases [4]. Whereas, similarity measures are possible to design for all kinds of fuzzy membership function pair if the similarity measures are designed by distance measure. With those designed similarity measure reliable data selection problem has been solved [7].

Mutual information analysis can be done by using FRIM or similarity measure between fuzzy sets  $A$  and  $B$ . Ding et al proposed the relative information which is based on the fuzzy entropy. Considered fuzzy entropy can be defined by the DM between two fuzzy sets. Furthermore, relation between fuzzy entropy and similarity measure has also studied [7], and counter meaning of similarity measure was defined by dissimilarity measure, in which dissimilarity measure was

derived through similarity and vice versa [5]. Those relations give us the result that two measures can be obtained through counter measure designing. Fuzzy relative information measure has been considered by way of similarity measure. With this result FRIM was considered via similarity measure.

In the next chapter, fuzzy entropy and similarity measure are introduced to describe the quantization of information. Fuzzy entropies derived by Shannon and Zadeh are introduced and explained. In Chapter III, conventional mutual information measure is compared with the proposed similarity measure. Finally, conclusions are followed in Chapter IV.

## 2. Preliminaries of similarity measure

It is well known that the fuzzy entropy depicts the degree of fuzziness for a fuzzy set. De Luca and Termini defined information entropy about fuzzy set as following mapping:

$$H : \xi(X) \rightarrow R^+$$

$$A \mapsto H(A),$$

where  $\xi(X)$  is a set consisting of all fuzzy subsets of universe of discourse  $X$ , and  $A \in \xi(X)$ .

**Definition 2.1** [8] Fuzzy entropy  $H$  satisfies following four axioms, that is:

- (i)  $H(A) = 0$  if and only if  $\mu_A(x) = 0$  or  $1$ ,  $\forall x \in X$
- (ii)  $H(A)$  takes the maximum value if and only if  $\mu_A(x) = 1/2$ ,  $\forall x \in X$
- (iii) if  $A \pi B$  then  $H(A) \leq H(B)$ , where  $A \pi B$  means  $A$  is a sharp set  $B$ , i.e.  
 $0 \leq \mu_A(x) \leq \mu_B(x) \leq 1/2$ , for  $0 \leq \mu_B(x) \leq 1/2$   
 $1/2 \leq \mu_B(x) \leq \mu_A(x) \leq 1$ , for  $1/2 \leq \mu_B(x) \leq 1$
- (iv)  $H(A) = H(\bar{A})$ , where  $\bar{A}$  is complementary set of  $A$ .

Manuscript received Jun. 29, 2010; revised Aug. 17, 2010;  
accepted Sep. 1, 2010

\* Corresponding Author: Sanghyuk Lee

Zadeh proposed the concept of fuzzy entropy in 1968 [9]. However, his definition does not satisfy the four axioms of fuzzy entropy, it is only a kind of weighted Shannon information entropy. Kaufmann also proposed the information entropy of fuzzy set, however it has the problem when the membership degree is the same [10]. Besides these results, Kosko and Pal and Pal proposed fuzzy entropies which satisfying De Luca and Termini fuzzy entropy [8].

Next, similarity measure between two sets is defined in Definition 2.2 [1]. On the contrary the properties of Definition 2.1 similarity measure shows that the degree of closeness between two sets containing fuzzy sets or ordinary sets.

**Definition 2.2** For  $\forall A, B \in F(X)$  and  $\forall D \in P(X)$ , similarity measure has following four properties

$$(S1) s(A, B) = s(B, A), \quad \forall A, B \in F(X)$$

$$(S2) s(D, D^c) = 0, \quad \forall D \in P(X)$$

$$(S3) s(C, C) = \max_{A, B \in F} s(A, B), \quad \forall C \in F(X)$$

$$(S4) \quad \forall A, B, C \in F(X), \quad \text{if } A \subset B \subset C, \quad \text{then } s(A, B) \geq s(A, C) \text{ and } s(B, C) \geq s(A, C),$$

$F(X)$  and  $P(X)$  denote fuzzy set and ordinary set, respectively.

These two definitions 2.1 and 2.2 reveal counter meaning each other, and their summation represent total information of dissimilarity and similarity measure respectively [5].

## 2.1 Illustrations of Fuzzy Entropies and Similarity measures

There are many fuzzy entropies satisfying Definition 2.1, following entropies are satisfying four axioms of Definition 2.1, and the proofs are found in our previous results [5, 6].

- Entropy of fuzzy data set with respect to the corresponding ordinary set can be designed using distance measure.

$$e(A, A_{near}) = d(A \cap A_{near}, [1]_X) + d(A \cup A_{near}, [0]_X) - 1$$

$$e(A, A_{near}) = d(A \cap A_{near}^c, [0]_X) + d(A \cup A_{near}^c, [1]_X)$$

$$e(A, A_{near}) = 1 - d(A \cap A_{near}, [0]_X) - d(A \cup A_{near}, [1]_X)$$

Where, crisp set  $A_{near}$  represents the crisp set ‘‘near’’ to the fuzzy set  $A$ .  $A_{near}$  is referred by variable as  $0 \leq near \leq 1$ . For example, the value of crisp set  $A_{0.5}$  represent one when  $\mu_A(x) \geq 0.5$ , and it is zero when  $\mu_A(x) \leq 0.5$ . Above fuzzy entropies are represent the degree of uncertainty between fuzzy set and corresponding deterministic ordinary set  $A_{near}$ .

Basically, conventional fuzzy entropy represent the DM between set  $A$  and  $A_{near}$ .

- Next, similarity measures between two data sets are also followed.

$$s(A, B) = d(A \cap B, [0]_X) + d(A \cup B, [1]_X)$$

$$s(A, B) = 1 - d(A \cap B^c, [0]_X) - d(A \cup B^c, [1]_X)$$

$$s(A, B) = 2 - d(A \cap B, [1]_X) - d(A \cup B, [0]_X)$$

where  $A \cap B$  and  $A \cup B$  are expressed the minimum and maximum value, expressions are commonly used in fuzzy set theory. Hence,  $(A \cap B)(x) = \min(A(x), B(x))$  and  $(A \cup B)(x) = \max(A(x), B(x))$ , respectively. The distance is

$$\text{defined by } d(A \cap B) = \frac{1}{n} \sum_{i=1}^n |\mu_A(x_i) - \mu_B(x_i)|. \quad [1]_X \text{ and } [0]_X$$

satisfy one and zero for the universe of discourse, respectively. Furthermore proofs of the conventional similarity measure can be also found in previous results [5].

Equations of fuzzy entropy and similarity can be also explained by graphical point of view. Fuzzy entropy means the degree of uncertainty or the dissimilarity between two data sets, fuzzy set and corresponding ordinary set generally. Hence, it can be design through many ways satisfying Definition 2.1. Similarity measure represents the degree of similarity between all kinds of data sets. Fuzzy entropy and similarity can be explained by graphical illustration in Fig. 1. From Fig. 1 shaded area represent the common information of two fuzzy sets with membership functions. Hence, regions  $C$  and  $D$  satisfy the definition of similarity measure. Except region of  $C$  and  $D$  satisfy the dissimilarity between two data sets. Therefore, it is denoted by fuzzy entropy or dissimilarity measure. By Fig. 1 the relation between similarity and dissimilarity has been emphasized in our previous result [5].



Fig. 1 Gaussian type two membership functions

In which the total information of two fuzzy set membership functions are represented by the summation of results similarity and dissimilarity measure. Non-convex fuzzy member are uncommon for the fuzzy set theory. However, non-convex fuzzy membership functions same results were also obtained [11].

## 2.2 Fuzzy Relative Information Measure

Relative measure was introduced by Ding et al [12]. Which means ‘‘the influence degree of the fuzzy set  $A$  to fuzzy set  $B$ ’’, and vice versa. They have described the fuzzy relative information measure of  $B$  to  $A$ ,  $R[A, B]$ .

$$R[A, B] = \frac{H(A \cap B)}{H(A)} = \frac{H(A) - H(A|B)}{H(A)} \quad (1)$$

is represented by an influence degree of the fuzzy set  $A$  to the fuzzy set  $B$ . Where,  $H(A)$  represents the entropy function based on the Shannon function:

$$H(A) = -\frac{1}{n} \sum_{x \in X^+} [\mu_A(x) \log \mu_A(x) + (1 - \mu_A(x)) \log(1 - \mu_A(x))] \quad \text{and}$$

$$H(A \cap B) = -\frac{1}{n} \sum_{x \in X^+} [\mu_B(x) \log \mu_B(x) + (1 - \mu_B(x)) \log(1 - \mu_B(x))] - \frac{1}{n} \sum_{x \in X^-} [\mu_A(x) \log \mu_A(x) + (1 - \mu_A(x)) \log(1 - \mu_A(x))]$$

Where,  $X^+ = \{x \mid x \in X, \mu_A(x) \geq \mu_B(x)\}$  and  $X^- = \{x \mid x \in X, \mu_A(x) < \mu_B(x)\}$  are satisfied respectively.  $R[A, B]$  satisfies the ratio between entropies of  $A \cap B$  and  $A$ . Hence, virtual ordinary set corresponding to the fuzzy set has to be needed.

### 3. Fuzzy Relative Information with Similarity Measure

Fuzzy relative information was analyzed through fuzzy entropy [12]. Fuzzy entropy is explained by comparing fuzzy set with respect to the corresponding ordinary set. In order to organize the relative information measure, virtual ordinary set has to be readied for entropy calculation. However, similarity measure can provide direct calculation between two fuzzy membership functions. Hence, relative information measure design with similarity measure can be efficient to minimize calculation time and decrease the design complexity.

#### 3.1 Characteristics of Relative Information Measure

Definition of relative information has not been formulated by researchers. In [12], they just proposed fuzzy relative information measure  $R[A, B]$  as the fuzzy relative information measure of  $B$  to  $A$ . Hence, definition of fuzzy relative information measure will be presented through analyzing the definition of  $R[A, B]$ .

**Proposition 3.1** Fuzzy relative information measure  $R[A, B]$  satisfies following properties:

- (i)  $R[A, B] = 0$  if and only if there is no intersection between  $A$  and  $B$ , or  $A, B$  are ordinary sets.
- (ii)  $R[A, B] = R[B, A]$  if and only if  $H(A) = H(B)$ .
- (iii)  $R[A, B]$  takes maximum value and  $R[A, B] \geq R[B, A]$  if and only if  $A$  is contained in  $B$ , i.e.  $\mu_A(x) \leq \mu_B(x)$  for  $\forall x \in X$ .
- (iv) If  $A \subset B \subset C$ , then  $R(B, A) \geq R(C, A)$  and  $R(A, B) = R(A, C) = R(B, C)$ .

Liu insisted that entropy can be calculated from the similarity measure and dissimilarity measure, which is denoted by  $s + d = 1$  [ ]. With this concept relative information measure can be designed via similarity measure. By the

definition of entropy for certain fact,  $H(A \cap B)$  and  $H(A)$  satisfy  $H((A \cap B), (A \cap B)_{near})$  and  $H(A, A_{near})$ , respectively. Where,  $(A \cap B)_{near}$  satisfies the same definition of  $A_{near}$ . Roughly, it can be satisfied that

$$R[A, B] = \frac{1 - s((A \cap B), (A \cap B)_{near})}{1 - s(A, A_{near})} \quad (2)$$

Where,  $s((A \cap B), (A \cap B)_{near}) = 1 - H((A \cap B), (A \cap B)_{near})$  and  $s(A, A_{near}) = 1 - H(A, A_{near})$ .

This measure also satisfies Proposition 3.1. Next, another relative information measure satisfying Proposition 3.1 without virtual ordinary sets  $(A \cap B)_{near}$  and  $A_{near}$  is considered.

#### 3.2 Fuzzy Entropy and Similarity Measure

Fuzzy relative information characteristic which satisfying Proposition 3.1 was proposed through entropy of fuzzy set  $A$  and  $B$ ,  $R[A, B]$ . With consideration the structure of similarity measure relative information measure satisfies following formation.

$$R[A, B] = \frac{s((A \cap B), (A \cap B)^C)}{s(A, A^C)} \quad (3)$$

By considering the characteristics of Proposition 3.1, (i) and(ii) are clear. (iii) also satisfies because  $A \cap B = A$  if  $\mu_A(x) \leq \mu_B(x)$ , furthermore  $R[A, B] = 1$  and  $R[B, A] \leq 1$  are satisfied naturally. Finally, (iv) is followed with the fact of (iii). In proposed (3), similarity measure can be replaced by explicit formulations of Chapter II.

- Next, computation of influence degree of fuzzy set to another fuzzy set is carried out through data selection problem. Illustration of data selection from a universal set has been done as follows:
  - (1) Selection of 5 students out of 65 students
  - (2) Selection trials are independent each other
  - (3) Selection is done randomly

Student point distribution has to be satisfied Gaussian type naturally. Table 1 shows the 65 students' point, whose mean is 52.7 and the standard deviation is 14.49.

Table1. Point list of 65 students

65 students points	82, 81.5, 76, 75, 75, 68, 67, 65.5, 65, 64.5, 64, 63.5, 63, 63, 62.5, 62, 61, 61, 60, 60, 60, 59, 59, 59, 58, 58, 58, 57.5, 57.5, 57, 56.8, 56, 55.5, 54, 53.5, 52.5, 52.5, 52.5, 52.5, 52, 51, 51, 49.5, 48, 47.5, 46.5, 46, 45.5, 45, 45, 44, 43, 41.5, 41, 40, 37, 37, 36, 33.5, 32, 31, 27, 26.5, 21, 0
	Mean :52.7, standard deviation : 14.5

With data of Table 1 data distribution is illustrated in Fig. 2. Data distribution can be also considered as the fuzzy set with

membership function,

$$A = \{ \langle x, \mu_{A_{middle}}(x) \rangle \mid x \in X, 0 \leq \mu_{A_{middle}}(x) \leq 1 \} .$$

$\mu_{A_{middle}}(x)$  is the middle grade student membership function.

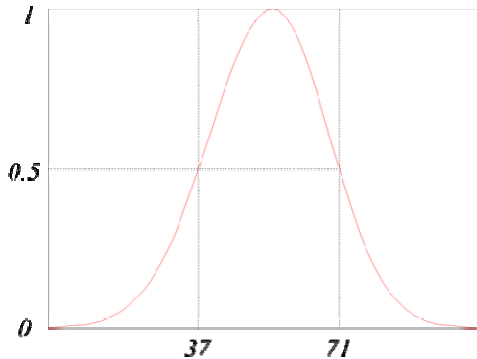


Fig. 2 Point distribution consideration as membership function

The average-level students have grades of B and C whose points are between 37 and 71. Every student who is contained in those areas can be called middle level or average level student, heuristically. However, the problem of how much they are in the middle level is a more delicate and philosophical problem. In this example computation of the fuzzy entropy and similarity measure represent the uncertainty and certainty with respect to the ordinary data set, i.e, grades B and C. By heuristic approach the second trial is the best among experiments, however computation of fuzzy entropy show the somewhat different result [7]. Test results are summarized in Table 2.

Table 2. Fuzzy entropy and Similarity of Sample data

	Sample	Membership value	Fuzzy entropy	Similarity value
Test 1	25	0.161	0.323	0.161
	44	0.835	0.329	0.835
	54	0.996	0.008	0.996
	61	0.849	0.302	0.849
	80	0.170	0.340	0.170
Average	52.8	0.600	0.260	0.6021
Test 2	50	0.983	0.034	0.983
	52	0.999	0.002	0.999
	55	0.987	0.025	0.987
	57	0.957	0.086	0.957
	59	0.990	0.180	0.990
Average	54.6	0.980	0.066	0.9832
Test 3	43	0.800	0.400	0.800
	52	0.999	0.002	0.999
	54	0.996	0.008	0.996
	55	0.987	0.025	0.987
	69	0.532	0.937	0.532
Average	54.6	0.860	0.275	0.8628
Test 4	12	0.019	0.039	0.019
	46	0.899	0.203	0.899
	53	1.000	0.000	1.000

	55	0.987	0.025	0.987
	91	0.031	0.016	0.031
Average	51.4	0.590	0.066	0.5872

By the meaning of fuzzy entropy, the entropy value approaches zero, the student group has a higher tendency toward B and C grade. The average entropy values of the groups obtained are 0.260, 0.066, 0.275, and 0.066. The fuzzy entropy results indicate that the 2nd and 4th trials are the nearest average level. Is it really certain? From the statistical point of view, the mean values of the trials are 52.8, 54.6, 54.6, and 51.4, respectively. The statistical results showed that the sample means of each case is similar to the total average; however, the 1st trial illustrates the nearest value to the mean.

Table 3. Difference mean and entropy of 4 trials

	Difference with mean	Fuzzy entropy
Test 1	0.1	0.260
Test 2	1.9	0.066
Test 3	1.9	0.275
Test 4	1.3	0.066

With the results, the similarity measure 0.602, 0.983, 0.863 and 0.587 are computed, respectively. The 2nd trial has the highest similarity value among 4 trials, hence it can be determined that Test 2 result is the nearest average level 5 students among the 4 times trials with only similarity measure. From this decision, with only similarity measure provides which trial is the most reliable data selection for this problem. To obtain same result fuzzy entropy calculation is needed more statistical information. Whereas compared to those results of fuzzy entropy, similarity measure has explicit advantage for reliable data selecting.

### 3.3 Analysis by Relative Information

Proposed relative information measure (3), denominator is the same for all trials. Hence, numerator comparisons are followed. In  $s((A \cap B), (A \cap B)^C)$ , Fuzzy set  $A$  is considered as the middle level fuzzy set, and its membership function satisfies Fig. 2. Whereas test data are also considered by fuzzy set  $B$ . Fig. 4 shows two membership functions between continuous and discrete cases.

Table 4. Computation of  $A \cap B$  and  $(A \cap B)^C$

	Sample	Membership value	$A \cap B$	$(A \cap B)^C$
Test 1	25	0.161	0.161	0.839
	44	0.835	0.835	0.165
	54	0.996	0.996	0.004
	61	0.849	0.849	0.151
	80	0.170	0.170	0.830
Test 2	50	0.983	0.983	0.017
	52	0.999	0.999	0.001

	55	0.987	0.987	0.013
	57	0.957	0.957	0.043
	59	0.990	0.990	0.010
Test 3	43	0.800	0.800	0.200
	52	0.999	0.999	0.001
	54	0.996	0.996	0.004
	55	0.987	0.987	0.013
	69	0.532	0.532	0.468
Test 4	12	0.019	0.019	0.981
	46	0.899	0.899	0.101
	53	1.000	1.000	0.000
	55	0.987	0.987	0.013
	91	0.031	0.031	0.969

Computation of  $s((A \cap B), (A \cap B)^c)$  has been followed with the similarity measure in Chapter II.

Table 5. Similarity computation

	$s((A \cap B), (A \cap B)^c)$
Test 1	0.2604
Test 2	0.0336
Test 3	0.2744
Test 4	0.0656

Values of Table 5 indicates “the influence degree of the fuzzy set  $A$  to fuzzy set  $B$ ”, i.e, “the influence degree of the middle level set to selection data”. Hence, if the value goes to zero selection is similar to the considering set. Actually, Test 2 is very similar to the middle level fuzzy set.

#### 4. Conclusions

For information data groups, each datum or data set can be represented by uncertainty or certainty for fixed numerical values. Furthermore, it also has a correlation between the degree of similarity and dissimilarity, these values are evaluated by fuzzy entropy and similarity measure. First, fuzzy entropy and similarity are introduced, and discussed their meaning and application. Fuzzy relative information measure has a role to represent the influence degree of fuzzy set to another fuzzy set. Measure was proposed by Ding et al, which was constructed through fuzzy entropy. With similarity measure, dual meaning of fuzzy entropy, another relative information measure is proposed, and characteristics is also proved. Data section problem was applied to verify the usefulness. Conventional results with fuzzy entropy and similarity measure are also compared. By simple calculation proposed relative information measure has its own properness to analyze relation between two data sets.

#### Acknowledgment

This work was supported by 2nd BK21 Program, which is

funded by KRF (Korea Research Foundation).

This work was supported by Priority Research Centers Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2010-0020163)

#### References

- [1] L. Xuecheng, “Entropy, distance measure and similarity measure of fuzzy sets and their relations”, *Fuzzy Sets and Systems*, vol. 52, pp. 305-318, 1992. doi: 10.1016/0165-0114(92)90239-Z
- [2] D. Bhandari and N.R. Pal, “Some new information measure of fuzzy sets”, *Inform. Sci.* vol. 67, pp. 209–228, 1993. doi: 10.1016/0020-0255(93)90073-U.
- [3] Ghosh, “Use of fuzziness measure in layered networks for object extraction: a generalization”, *Fuzzy Sets and Systems*, vol. 72, pp. 331–348,1995. doi: 10.1016/0165-0114(94)00291-E.
- [4] S.J. Chen and S.M. Chen. Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *IEEE Trans. on Fuzzy Systems*, vol. 11, no. 1, pp.45-56, 2003. doi: 10.1109/TFUZZ.2002.806316.
- [5] S.H. Lee, W. Pedrycz, and Gyoyong Sohn, “Design of Similarity and Dissimilarity Measures for Fuzzy Sets on the Basis of Distance Measure”, *International Journal of Fuzzy Systems*, vol. 11, pp. 67-72, 2009.
- [6] S.H. Lee, K.H.Ryu, G.Y. Sohn, “Study on Entropy and Similarity Measure for Fuzzy Set”, *IEICE Trans. Inf. & Syst.*, vol. E92-D, pp. 1783-1786, 2009.
- [7] S.H. Lee, Y.T. Kim, S.P. Cheon, and S.S. Kim, “Reliable data selection with fuzzy entropy”, *LNAI*, vol. 3613, pp. 203–212, 2005.
- [8] DeLuca and S. Termini, “Adefinition of nonprobablistic entropy in the setting of fuzzy sets theory”, *Information and Control*, vol. 20, no. 3, pp. 301-312, 1972. doi:10.1016/S0019-9958(72)90199-4.
- [9] L.A. Zadeh, “Probabilistic measures of fuzzy events”, *Journal of Mathematical Analysis and Applications*, vol. 23, no. 10, pp. 421-727, 1968
- [10] Kaufmann(ed.), *Introduction to the Theory of Fuzzy Subsets-Fundamental Theoretical Elements*, Academic Press, New York, 1975.
- [11] S.H. Lee, S. J. Kim, N. Y. Jang, “Design of Fuzzy Entropy for Non Convex Membership Function”, *CCIS*, vol. 15, pp. 55–60, 2008. doi:10.1007/978-3-540-85930-7.
- [12] S.F. Ding, S.H. Xia, F.X. Jin, Z.Z. Shis, “Novel fuzzy information proximity measures”, *Journal of Information Science*, vol. 33, no. 6, 678-685, 2007. doi: 10.1177/0165551507076332.



**Hong-mei Wang** received the B.S degree in Automation Control Engineering at Qingdao University, China, in 2006. Now she is a Ph.D candidate in School of Mechatronics Engineering, Changwon National University, Korea. Her research interests include in the area of designing system applied for wireless communication modem and various systems required advanced digital signal processing



**Sanghyuk Lee** received the B.S. in EE from Chungbuk National University, in 1988, M.S. and Ph.D. degrees in EE from Seoul National University, in 1991 and 1998, respectively. Dr. Lee served as a Research Fellow from 1996 to 1999 in HOW Co.. He is currently a Research Professor in Institute for Information and Electronics Research in Inha University as a Research Professor. His research interests include fuzzy theory, game theory, controller design for linear and nonlinear systems.