한국지능시스템학회 논문지 2009, Vol. 19, No. 6, pp. 809-813

최강력 검정을 위한 퍼지 포아송 가설의 검정

Fuzzy Hypothesis Test by Poisson Test for Most Powerful Test

강만기* · 서현아**

Man Ki Kang* and Hyun A Seo**

* Dept. of Data Information Science, Dong-Eui University, Busan, 614-714, Korea

** Dept. of Information Statistics, Graduate School, Dong-Eui University, Busan, 614-714, Korea

Abstract

We want to show that the construct of best fuzzy tests for certain fuzzy situations of Poisson distribution. Due to Neyman and Pearson theorem, if we have θ_0 and θ_1 be distinct fuzzy values of $\Omega=\{\theta:\theta=\theta_0,\theta_1\}$ such that $L(\theta_0:X)/L(\theta_1:X)<k$, then k is a fuzzy number. For each fuzzy random samples point $X\subset C$, we have most power test for fuzzy critical region C by agreement index.

Key Words : optimal tests of hypotheses, most powerful test, best critical region, likelihood ratio.

1. Preliminaries

We propose some properties of fuzzy Poisson test using most powerful test by agreement index and obtained from the fuzzy samples, the negation of the assertion is taken to be the fuzzy null hypothesis H_{f0} and the assertion itself is taken to be the fuzzy alternative hypothesis H_{f1}([1],[6]).

Kang, Choi and Lee[2] and Kang, Lee, and Han[4] defined fuzzy hypotheses membership function also they found the agreement index by area for fuzzy hypotheses membership function and membership function of fuzzy critical region, thus they obtained the results by the grade for judgement to acceptance or rejection for the fuzzy hypotheses.

First we define fuzzy probability for repeatedly observed data with alteration error term. In chapter 3, we introduction some properties of fuzzy hypotheses test by agreement index. For few mean, we show that a Poisson function of performance for a fuzzy hypothesis test and drawing conclusions from the most powerful test in chapter 4.

We considered the fuzzy hypothesis

$$H_{f0}:\theta\simeq\theta_0 \text{ or } H_{f0}:\theta\prec\theta_0,\ \theta\in\Omega \tag{1.1}$$

where "$\simeq$" is similarity and "$\prec$" is less than or similarity, and constructed by a set

$$\{(H_{f0}(\theta),H_{f1}(\theta))|\theta\in\Omega\} \tag{1.2}$$

with membership function $m_{H_f}(\theta)$ where Ω is parameter space and the fuzzy hypotheses H_{f0} is referred to as the null fuzzy hypothesis while H_{f1} is referred to as the alternative fuzzy hypothesis.

A sample fuzzy number A in R is said to be convex if for any real numbers $x,y,z\in R$ with $x\le y\le z$,

$$m_A(y)\geqq m_A(x)\wedge m_A(z) \tag{1.3}$$

with $\wedge$ standing for minimum.

A sample fuzzy number A is called normal if the following holds

$$\bigvee_x m_A(x)=1. \tag{1.4}$$

An $\delta-$level set of a sample fuzzy number A is a set denoted by $[A]^\delta$ and is defined by

$$[A]^\delta=\{x|\,m_A(x)\geqq\delta,\,0<\delta\leqq 1\} \tag{1.5}$$

An $\delta-$level set of sample fuzzy number A is a convex fuzzy set which is a closed bounded interval denoted by $[A]^\delta=[A_l^\delta,A_r^\delta]$.

2. Fuzzy probability for random experiment

The concept of probability is relevant to experiments that have same what uncertain outcomes. Thus uncertain outcomes is fuzzy concepts.

We denote by $x=x(s)$ the possible outcome of an fuzzy random experiment of subject s and will be call the fuzzy sample point.

Definition 2.1. The sample of an fuzzy random experi-

접수일자 : 2009년 9월 10일
완료일자 : 2009년 11월 30일

ment of subject s is a pair $(\Omega(s),\Phi)$, where

(1) $\Omega(s)$ is the set of all possible outcomes of the fuzzy random experiment of subject s

(2) Φ is a σ-field of subjects of $\Omega(s)$.

Let Φ be the σ-field that is generated by such an event family, any set $A\in\Phi$ is called an event.

We must note that now we do not have any probability measure on sample space (Ω,S), next let us discuss how one can construct a probability on (Ω,S) by the sample of fuzzy random experiment $(\Omega(s),\Phi)$, satisfying the Kolmogorov axiom[7].

The fuzzy random experiment can be repeated under identical fuzzy condition or common condition by one and the same subject s, for $i=1,2,\cdots$; we denote the k-th trial of subject s by $x_k = x_k(s)$. For each $A\in S$, let $m_A(x)$ be given by

$$0 \leqq m_A(x) \leqq 1 \tag{2.1}$$

If we have a fuzzy random experiment with repeated as n times devide as $n=n_1+n_2$, $n_i>0$, $i=1,2$, the fuzzy number relative frequency of event A are

$$A_1^{(\delta)} = [A_{1l},A_{2r}]^{(\delta)} = \frac{1}{n_1}\sum_{k=1}^{n_1}\{x_k|m_A(x_k) \geqq \delta\}$$
$$= \frac{1}{n_1}\sum_{k=1}^{n_1}\{x_k|m_A(x_k(s) \geqq \delta\}. \tag{2.2}$$

$$A_2^{(\delta)} = [A_{2l},A_{2r}]^{(\delta)} = \frac{1}{n_2}\sum_{k=n_1+1}^{n}\{x_k|m_A(x_k) \geqq \delta\}$$
$$= \frac{1}{n_2}\sum_{k=n_1+1}^{n}\{x_k|m_A(x_k(s) \geqq \delta\} \tag{2.3}$$

Thus we have

$A_l = Min_i\{A_i^{(\delta)}\}$, $A_r = Max_i\{A_i^{(\delta)}\}$ $i=1,2$

and $A_{c=}\frac{1}{n}(A_{1l}{*}n_1 + A_{2r}{*}n_2)$ if $A_{1l} \leqq A_{2l}$

and $A_{1r} \leqq A_{2r}$

or $A_{c=}\frac{1}{n}(A_{2l}{*}n_2 + A_{1r}{*}n_1)$ if $A_{1l} \geqq A_{2l}$

and $A_{1r} \geqq A_{2r}$ elsewhere

$$A_{c=}\frac{1}{2n}(n_1{*}(A_{1l}+A_{1r}) + {*}n_2(A_{1l}+A_{2r})). \tag{2.4}$$

From (2.2)-(2.4), we have fuzzy number probability as

$$P(A,n,s) = A = [A_l, A_c, A_r] \tag{2.5}$$

with membership function $m_A(x_k)$.

Thus we have following two propositions.

Proposition 2.1 For any fuzzy random experiment, let (Ω,S) be its common sample space. Then in a sequence of fuzzy random experiment of subject $s(\in S)$, the probability $P(A,n,s)$, $n=1,2,\cdots$, has all the properties as follows. For each $A\in S$,b and for each $n\geq 1$;

(1) $0 \leq P(A,n,s) \leq 1$ (2.6)

(2) $P(\Omega,n,s) = 1$ (2.7)

(3) If $A_i \in S$ and $A_i \cap A_j = \phi$, $i\neq j$; $i,j=1,2,\cdots$, let

$$A = \bigcup_{i=1}^{\infty} A_i, \text{ then } P(A,n,s) = \sum_{i=1}^{\infty} P(A_i,n,s). \tag{2.8}$$

Proof. For each $1\leq k\leq n$, if we have fuzzy samples $\omega_k(s)$ then $m_A(x_k(s))$ is fixed a set functions satisfying $0 \leqq m_A(x_k(s)) \leqq 1$ for each $A\in S$ and $m_\Omega(x_k(s)) = 1$ is clear.

For any sequence $\{A_i\}$ of disjoint sets of S, let $A = \cup_{k=1}^{\infty} A_i$, if $x_k(s)\in A$, $1\leq k\leq n$ then

$$P(A,n,s) = \frac{1}{n}\sum_{k=1}^{n} m_A(x_k(s)) = \frac{1}{n}\sum_{k=1}^{n} m_{\bigcup_{i=1}^{\infty} A_k}(x_k(s))$$
$$= \frac{1}{n}\sum_{k=1}^{n} Max\{m_{A_1}(x_k(s)),\, m_{A_2}(x_k(s)),\cdots\}$$
$$= \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{\infty} m_{A_i}(x_k(s)) = \sum_{i=1}^{\infty} P(A_i,n,s).$$

It follows $0\leq P(A,n,s)\leq 1$ for each $A\in S$,

$$P(\Omega,n,s) = 1 \quad \text{and } P(A,n,s) = \sum_{i=1}^{\infty} P(A_i,n,s).$$

Proposition 2.2. We have that

(1) $P(A^c,n,s) = 1 - P(A,n,s)$. (2.9)

(2) $P(\phi,n,s) = 0$. (2.10)

(3) If $A_1, A_2 \in \Phi$ then

$$P(A_1\cup A_2,n,s) = P(A_1,n,s) + P(A_2,n,s) - P(A_1\cap A_2,n,s). \tag{2.11}$$

Proof. Since

$$P(A^c,n,s) = \frac{1}{n}\sum_{k=1}^{n} m_{A^c}(x_k) = \frac{1}{n}\sum_{i=1}^{n}(1 - m_A(x_k(s)))$$
$$= \frac{1}{n}\sum_{i=1}^{n} 1 - \frac{1}{n}\sum_{i=1}^{n} m_A(x_k(s)) = 1 - \frac{1}{n}\sum_{i=1}^{n} m_A(x_k(s)).$$

So we have

$P(A^c,n,s) = 1 - P(A,n,s)$.

From $P(A_1\cup A_2,n,s) = \frac{1}{n}\sum_{k=1}^{n} m_{A_1\cup A_2}(x_k)$

$$= \frac{1}{n}\sum_{k=1}^{n} Max\{m_{A_1}(x_k), m_{A_2}(x_k)\}.$$

Put $S_1 = \{x_k \in S|\, m_{A_1}(x_k) \geq m_{A_2}(x_k)\}$,

by n_1 times for $A_1(x_k)$ and

$S_2 = \{x_k\in S\,|\, m_{A_1}(x_k) < m_{A_2}(x_k)\}$,

by n_2 times for $A_2(x_k)$ and $S = S_1\cup S_2$, $n_1+n_2 = n$.

Then

$$P(A_1 \cup A_2, n, s) = \frac{1}{n}\sum_{k=1}^{n_1} m_{A_1}(x_k|S_1) + \frac{1}{n}\sum_{k=1}^{n_2} m_{A_2}(x_k|S_2).$$

Since $P(A_1 \cap A_2, n, s) = \frac{1}{n}\sum_{k=1}^{n} m_{A_1 \cap A_2}(x_k)$

$$= \frac{1}{n}\sum_{k=1}^{n} Min\{m_{A_1}(x_k),\ m_{A_2}(x_k)\}$$

$$= \frac{1}{n}\sum_{k=1}^{n_2} m_{A_1}(x_k|S_2) + \frac{1}{n}\sum_{k=1}^{n_1} m_{A_2}(x_k|S_1).$$

Thus we have

$$P(A_1 \cup A_2, n, s) = P(A_1, n, s) + P(A_2, n, s) - P(A_1 \cap A_2, n, s)$$

So we have

1. For set $\Omega(s)$ of subject $s(\in S)$.
2. Any σ-field Φ in $\Omega(s)$.
3. Set function $P(A,n,s)$ is a normal measure on ϕ.

So every such triple $(\Omega(s), \Phi, P(A,n,s))$ will be call a probability space according to the viewpoint of modern probability theory.

3. Acceptance or rejection degree

Let X be a random variable by fuzzy random sample from sample space Ω. Let $\{P_\theta, \theta \in \Omega\}$ be a family of fuzzy probability distribution, where Θ is a parameter vector of Ω.

Choose a membership function $m_X(x)$ whose value is likely to best reflect the plausibility of the fuzzy hypothesis being tested.

Let us consider membership function $m_C(x)$ of critical region C, which we will call the agreement index of $m_X(x)$ which regard to$m_C(x)$([3],[5]).

Definition 3.1. Let a fuzzy membership function $m_X(x)$, $x \in R$, we consider another membership function $m_C(x)$, $x \in R$, which call the agreement index, the ratio being defined in the following way;

$$AGI(X, C) = \frac{area(m_X(x) \cap m_C(x))}{area(m_C(x))} \in [0,1]. \quad (3.1)$$

Definition 3.2. We define the maximum grade membership function of rejection or acceptance degree by agreement index for real-valued function R_δ by δ−level on ψ as

$$\Re_\delta(0) = \sup_\theta \left\{ \frac{\text{area } (\mathrm{m}_{X_\delta}(\theta) \cap \mathrm{m}_{C_\delta}(\theta))}{\text{area } \mathrm{m}_{C_\delta}(\theta)} \right\}, \quad (3.2)$$

$$\Re_\delta(1) = 1 - \Re_\delta(0) \quad (3.3)$$

for the fuzzy hypothesis testing.

In agreement index, we have the area by δ−level as:

$$area(m_C(x) \cap m_X(x)) = \int_{\delta_0}^{\delta_1} (C_r^{-1}(\delta) - X_l^{-1}(\delta))d\delta$$

$$area\ m_{C_\delta}(\theta) = \int_{\delta_0}^{1} (C_r^{-1}(\delta) - C_l^{-1}(\delta))d\delta \quad (3.4)$$

where C_r, C_l are right and left side line of $m_C(x)$, X_l is left side line of $m_X(x)$ and δ_0 is reliable degree and δ_1 is meeting point of $m_C(x)$ and $m_X(x)$.

Definition 3.3. We have acceptance region and rejection region for the fuzzy critical region C as Fig 3.1.

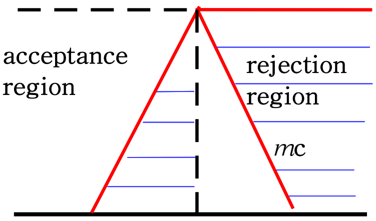
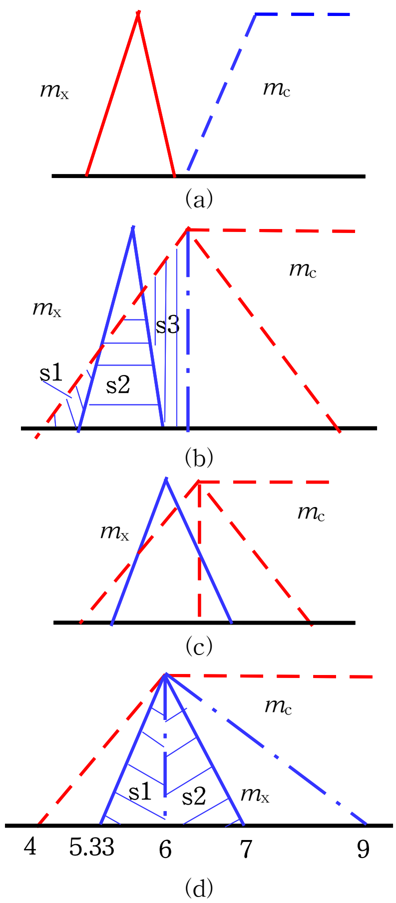


Fig 3.1 Acceptance and rejection region

For various kinds of X, we can reject the hypotheses by Definition 3.2 as [Fig 3.2].



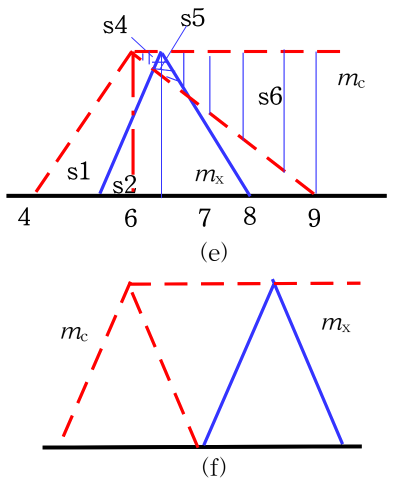


Fig 3.2 Various type of C by X

For example, in [Fig 3.2] (a) and [Fig 3.2] (f), we can clearly reject the hypothesis as degree $\Re_\delta(0)=0$ and $\Re_\delta(0)=1$.

For [Fig 3.2] (b), we have

$$\Re_\delta(0)=\frac{area(s_1)+area(s_2)}{area(s_1)+area(s_2)+area(s_3)}\times\frac{1}{2} \quad (3.5)$$

for left hand side area of center of fuzzy critical region C by fuzzy the statistics X.

In case of Fig 3.2 (d), we nave

$$\Re_\delta(0)=\frac{area(s_1)+srea(s_2)}{area(s_1)+srea(s_2)}\times\frac{1}{2}=0.5, \quad (3.6)$$

it's maintain an uncertain attitude for decision the hypotheses.

Also, we have

$$\Re_\delta(0)=\frac{area(s_1)+area(s_2)}{area(s_1)+area(s_2)}\times\frac{1}{2}$$
$$+\frac{area(s_4)+area(s_5)}{area(s_4)+area(s_5)+area(s_6)}\times\frac{1}{2} \quad (3.6)$$

in Fig 3.2 (e).

4. Most powerful test of fuzzy Poisson probability

We are interested in fuzzy random variable X which has fuzzy probability density function(p.d.f.) or probability mass function(p.m.f) $f(x:\theta)$ where $\theta\in\Omega$, Ω is parameter space. We assume that $\theta\simeq\theta_0$ or $\theta\simeq\theta_1$ where θ_0 and θ_1 are fuzzy subset of Ω.

We label the fuzzy hypothesis as

$$H_{f0}:\theta\simeq\theta_0 \text{ versus } H_{f1}:\theta\simeq\theta_1. \quad (4.1)$$

The test of H_{f0} versus H_{f1} is based on a fuzzy sample $X_1,X_2,\cdots,X_n$ from the distribution of $f(x:\theta)$.

A test of H_{f0} versus H_{f1} is based on a fuzzy subset C of sample space S. The fuzzy set C is called the fuzzy critical region and its corresponding decision degree rule is:

reject

$$H_{f0} \text{ by } \Re_\delta(0)(\text{accept } H_{f1} \text{ by } \Re_\delta(1)) \text{ if } X\subset C \quad (4.2)$$

or retain

$$H_{f0} \text{ by } \Re_\delta(1)(\text{reject } H_{f1} \text{ by } \Re_\delta(0)) \text{ if } X\subset C^c. \quad (4.3)$$

Note that a fuzzy test is defined by fuzzy critical region. Conversely a fuzzy critical region defines a fuzzy test.

The size of significance level of the test is denote by the probability of a type I error, i.e.,

$$\tilde{\alpha}=\max_{\theta\simeq\theta_0}P_\theta(X\subset C). \quad (4.4)$$

Also, the power function of a fuzzy test is given by

$$\gamma_C(\theta)=P_\theta(X\subset C);\ \theta\simeq\theta_1. \quad (4.5)$$

We will show the fuzzy significance level and fuzzy power function in next time.

In general, there will be a multiplicity of fuzzy subsets C' of the sample space such that $P_{\theta_0}(X\subset C')=\tilde{\alpha}$. Suppose that there is one of these fuzzy subset C, such that when H_{f1} is true, the power of the test associated with C is greaer than the power of the test associated with each other C'. Then C is defined as a best fuzzy critical region of size α for testing H_{f0} against H_{f1}.

Well known Neyman and Pearson theorem, if we have fuzzy random sample $X_1,X_2,\cdots,X_n$ from $f(x:\theta)$ then the joint p.d.f. of $X_1,X_2,\cdots,X_n$ is

$$L(\theta;x_1,x_2,\cdots,x_n)=f(x_1,\theta)f(x_2,\theta)\cdots f(x_n,\theta). \quad (4.6)$$

Let θ_0 and θ_1 be distinct value of θ, and k be a positive fuzzy number. Let C to be the subset of sample space which satisfy

$$\frac{L(\theta_0;x_1,x_2,\cdots,x_n)}{L(\theta_1;x_1,x_2,\cdots,x_n)}<k,\quad k>0, \quad (4.7)$$

for each point $(x_1,x_2,\cdots,x_n)\in C$

then, in accordance with the theorem, C will be a best critical region of size α for testing the fuzzy simple hypothesis.

This inequality can be can frequently expressed in form $X=u(x_1,x_2,\cdots,x_n;\theta_0,\theta_1)<c$, where c is a constant. Since θ_0 and θ_1 are given constants, $X=u(x_1,x_2,\cdots,x_n;\theta_0,\theta_1)$ is fuzzy statistics; and if the p.d.f. of this statistics can be found when H_{fo} is true, then the significance level of fuzzy test of H_{f0}.

An illustrative example follows.

Let $X_1, \ldots, X_n$ denote fuzzy random samples from a distribution which has a pmf $f(x;pt)$ of Poisson distribution.

If we have probability $\widetilde{p_0}=[0.09,\ 0.11]$ and $\widetilde{p_1}=[0.19,\ 0.21]$ per unit interval by any random experiment for level $\delta=0$ then we have mean $\theta_0=\widetilde{p_0}t=[0.9,\ 1.1]$ and $\theta_1=\widetilde{p_1}t=[1.9,\ 2.1]$ for time $t=10$.

It is desired to test the fuzzy hypothesis

$$H_{fo}: \theta \simeq \theta_0$$

against the alternative simple hypothesis

$$H_{f1}: \theta \simeq \theta_1.$$

Here,

$$\frac{L(\theta_0;\, x_1, x_2, \cdots, x_n)}{L(\theta_1;\, x_1, x_2, \cdots, x_n)} = \frac{\prod_1^n e^{-\theta_0}\theta_0^{x_i}/x_i!}{\prod_1^n e^{-\theta_1}\theta_1^{x_i}/x_i!} \tag{4.8}$$

$$= e^{n(\theta_0-\theta_1)}\left(\frac{\theta_0}{\theta_1}\right)^{\sum_1^n x_i} < \widetilde{k}\ ,\ \tilde{k} > 0 \tag{4.9}$$

If $k > 0$, from the set of points $(x_1, x_2, \ldots, x_n)$ we have

$$\left(\sum_1^n x_i\right)\ln\frac{\theta_0}{\theta_1} + n(\theta_1-\theta_0) < \ln\tilde{k} = C \tag{4.10}$$

since $\ln\dfrac{\theta_0}{\theta_1} < 0$,

$$X = \sum_1^n x_i > \frac{\ln\tilde{k} - n(\theta_1-\theta_0)}{\ln\dfrac{\theta_0}{\theta_1}} = C \tag{4.11}$$

is a best fuzzy critical region C. Consider the case of $k=[0.86,\ 9.0,\ 0.94]$ and $n=4$, we have the best fuzzy critical region $C=[3.9,\ 6.0,\ 9.0]$.

If we have random samples

$x_1=[1.3, 1.4, 1.6]$, $x_2=[1.4, 1.5, 1.6]$,

$x_3=[1.0, 1.1, 1.2]$, $x_4=[0.8, 1.0, 1.1]$

then $X=[4.5, 5.0, 5.5]$, the reject degree is $\Re_\delta(0)=0.36$ by [Fig 3.2] (b).

The case of data $X=[5.3, 6.0, 7.0]$ then reject degree is $\Re_\delta(0)=0.5$ as [Fig 3.2] (d).

Finally, if we have $X=[5.5, 6.5, 8.0]$ then reject degree is $\Re_\delta(0)=0.78$ by [Fig 3.2] (e).

References

[1] P. X. Gizegorzewski, Testing Hypotheses with vague data, *Fuzzy Sets and Systems*. 112 , pp.501-510, 2000.

[2] M. K. Kang, G. T. Choi and C. E. Lee, On Statistical Test for Fuzzy Hypotheses with Fuzzy Data, *Proceeding of Korea Fuzzy Logic and Intelligent System Society Fall Conference*, Vol. 10, Num. 2, 2000.

[3] M. K. Kang, G. T. Choi and S. I .Han, A Bayesian Fuzzy Hypotheses testing with Loss Function, *Proceeding of Korea Fuzzy Logic and Intelligent Systems Society Fall Conference,* Vol 13, Num. 2, 2002.

[4] M. K. Kang, C. E. Lee and S. I. Han, Fuzzy Hypotheses Testing for Hybrid Numbers by Agreement Index, *Far East Journal of Theoretical Statistics,* 10(1), 2003.

[5] M. K. Kang, Y. R. Park, Fuzzy Binomial Proportion Test by Agreement Index, *Journal of Korean Institute of Intelligent Systems,* Vol 19, Num. 1, 2009.

[6] W. Trutschnig, A strong consistency result for fuzzy relative frequencies interpreted as estimator for the fuzzy-value probability, *Fuzzy Sets and Systems*, 159, pp 259-269, 2008.

[7] Z. Xia, Fuzzy Probability System: fuzzy probability space(1), *Fuzzy Sets and Systems*, 120, pp.469-486, 2001.

저 자 소 개

강만기(Man Ki Kang)
한국지능시스템학회 논문지 Vol.15 No.2, Vol.17, No.5 참조

서현아(Hyun A Seo)
현재 : 동의대학교 대학원 정보통계학과 박사과정

관심분야 : 퍼지데이터처리