한국지능시스템학회 논문지 2008, Vol. 18, No. 2, pp. 200-205

유사측도를 이용한 신뢰성 있는 데이터의 추출

Reliable Data Selection using Similarity Measure

류수록 · 이상혁*
Soorok Ryu and Sang-Hyuk Lee

경북대학교 산업응용수학과
*창원대학교 메카트로닉스공학부

요 약

데이터 분석을 위하여 데이터의 불확실성에 대한 측도로서 퍼지 집합에 대한 엔트로피를 소개하였고, 또한 데이터간의 유사도를 나타내는 유사측도를 구성하였다. 퍼지 소속 함수간의 유사측도는 거리측도를 이용하여 구성하였고, 제안한 유사측도를 증명을 통하여 확인하였다. 제안한 유사측도의 유용성을 확인하기 위하여 신뢰성 있는 데이터추출 예제에 적용하였다. 적용결과를 퍼지 엔트로피와 통계적 지식을 통하여 얻어진 이전의 결과와 비교하였다.

키워드 : 퍼지 엔트로피, 유사측도, 거리측도

Abstract

For data analysis, fuzzy entropy is introduced as the measure of fuzziness, similarity measure is also constructed to represent similarity between data. Similarity measure between fuzzy membership functions is constructed through distance measure, and the proposed similarity measure are proved. Application of proposed similarity measure to the example of reliable data selection is also carried out. Application results are compared with the previous results that is obtained through fuzzy entropy and statistical knowledge.

Key Words : Fuzzy entropy, Similarity measure, Distance measure

1. 서 론

신뢰성 있는 데이터의 추출은 시스템 모델링 또는 데이터 분석 등에 적용하기에 적절한 연구 분야 이다. 적절한 유사측도의 구성을 통하여 패턴인식, 데이터 클러스터링, 그리고 정보예측 등의 분야에 적용할 수 있다. 일반집합 또는 퍼지 집합에 대한 신뢰성 있는 데이터의 추출을 위한 연구는 데이터의 확실성 또는 불확실성을 계량적으로 표현하고자 하는 방향으로 진행되어 왔다[1-11]. 계량적 연산을 위한 연구는 데이터의 불확실성을 계산하는 퍼지 엔트로피 구성에 대한 부분과, 데이터간의 유사측도를 구성하는 부분으로 분류할 수 있다. 이중 퍼지 엔트로피의 구성은 퍼지 집합에 대하여 퍼지 엔트로피의 정의를 만족하는 형태로 구성할 수 있으며, 보편적으로 거리측도를 이용한 형태를 취하였다 [1-4]. 유사측도의 경우는, 퍼지 엔트로피와 보완적인 관계를 형성한다. 유사측도 역시 거리측도를 이용하면 보다 구체적인 형태로의 유사측도 구성이 용이해진다. 따라서 거리측도를 이용한 유사측도의 구성에 대한 연구가 제안 되어왔다[6]. 또한 소속 함수의 형태가 삼각형 또는 마름모꼴의 경우, 퍼지 넘버를 이용한 유사측도의 구성에 대한 연구도 진행되었다 [7-11]. 그러나 퍼지 넘버를 이용하면 소속 함수의 형태에 제한을 받아서 일반적인 퍼지집합의 경우 적용하기에 어려움이 따른다.

이렇게 구성된 퍼지 엔트로피를 이용하여 데이터의 불확실성을 계산함으로서 보다 신뢰성이 높은 데이터의 취득이 가능하였다[4]. 그러나 퍼지 엔트로피만을 이용할 경우, 퍼지 엔트로피의 정의 중 여집합에 대한 특성, $e(A)=e(A^c), \forall A \in F(X)$, 으로 인하여 퍼지 엔트로피뿐 아니라 추가적인 정보가 요구 되었다 [5]. 따라서 이러한 문제점을 극복하기 위한 방법으로 유사측도를 고려한다. 유사측도의 경우 퍼지 데이터 상호간의 정보만을 고려하고, 구성 또한 퍼지 엔트로피와 동일한 방법으로 거리측도를 이용하여 구성하였다. 두 퍼지 집합간의 확실성을 나타내는 겹쳐진 소속 함수의 면적에 비례하는 형태로 유사측도를 구성하였다. 이를 위하여 먼저, 두 개의 퍼지 집합에 대한 해석을 통하여 두 퍼지집합 사이의 관계를 규명한다. 완전히 일치하지 않는 두 개의 퍼지 소속 함수는 겹친 부분과 겹치지 않은 부분으로 나눌 수 있다. 겹치지 않은 부분에 대한 분석을 통하여 우리는 퍼지 엔트로피를 구성하였고 그 결과를 얻었다 [5].

우리는 본 논문에서, 겹친 부분의 면적에 비례하는 형태로 유사측도를 구성하였다. 구성된 유사측도 만으로 신뢰성 있는 데이터의 추출을 실시하면 퍼지 엔트로피를 이용한 경우와 비교하여 다른 지식의 고려함 없이 유사측도 만으로의 연산이 가능해졌다. 제안한 유사측도의 적절함은 유사측도의 정의를 만족함을 증명을 통하여 확인하였다. 그리고 이전 우

접수일자 : 2007년 5월 14일
완료일자 : 2008년 1월 29일
* 교신저자
본 연구는 2단계 BK21사업의 지원을 받았습니다.

리의 연구결과에 적용하였던 예제에 대하여 구성된 유사측도를 적용하여 유용성을 확인한다. 본 논문의 구성은 다음과 같다.

다음 장에서 우리는 퍼지 집합과 유사측도의 관계를 설명하였다. 이를 위하여 먼저 기존의 퍼지 엔트로피를 소개하고, 특성을 분석하였다. 퍼지 집합들 간의 확실성과 불확실성 관계를 설명하고 유사측도와의 관계를 알아보았다. 또한 신뢰성 있는 데이터의 추출을 위하여 퍼지집합의 특성파악을 바탕으로 유사측도를 구성하였다. 3장에서 제시된 예제는 중간성적의 학생 5명을 전체 65명의 클래스에서 추출하는 것이 목적이다. 이때 학생 추출은 무작위로 실시하며, 추출된 학생들의 점수가 얼마나 중간수준에 충실한가의 판단을 제안한 유사측도와 이전의 결과인 퍼지엔트로피 그리고 통계적 지식을 통하여 얻은 결과와 비교 하였다[5]. 그리고 비교, 검토가 이루어졌다. 마지막으로 4장에서 결론을 맺는다. 본 논문에서 사용된 용어는 Liu의 문헌에서 사용된 수식에 따른다 [1].

2. 퍼지 집합과 유사측도

본 장에서는 퍼지 집합간의 확실하거나 불확실한 정보를 분석하는 두 가지 방법에 대하여 소개한다. 알려진 대로 정보의 불확실성을 계량적으로 나타내기 위한 측도로서 엔트로피와 퍼지 집합간의 유사도를 표현하기위한 유사측도를 소개한다.

2.1 퍼지 엔트로피

엔트로피는 불확실성에 대한 측도로서 정보 이론, 시스템 해석, 열역학 등의 분야에서 널리 적용되어왔다. 발생확률에 따른 엔트로피는, 발생확률이 아주 낮거나 높으면 엔트로피는 0에 가까워지고 발생확률이 1/2의 경우, 최대의 엔트로피를 갖는 것으로 알려져 있다. 이렇게 어떤 사건의 발생확률에 대한 엔트로피의 개념을 확장하여, 퍼지 집합의 소속 함수에 대하여 엔트로피를 구성하는 연구가 최근 진행되어 왔다[1-4]. 먼저 소속 함수를 이용한 퍼지 엔트로피에 대한 정의를 다음과 같이 나타낸다.

정의 2.1 [1] 퍼지 집합 $F(X)$ 에 대하여 $e: F(X)\rightarrow R^{+}=[0,\infty)$ 같은 함수가 성립하고 다음과 같은 특징을 만족하면 e는 퍼지 엔트로피라고 정의 한다:

(E1) $e(D)=0,\ \forall D\in P(X)$
(E2) $e([1/2])=\max_{A\in F(X)}e(A)$
(E3) $e(A^{*})\le e(A)$, 여기서 A^{*}는 퍼지 집합 A의 뾰족함.
(E4) $e(A)=e(A^{c}),\ \forall A\in F(X)$.

여기서 X 는 전체집합이고, $[1/2]$ 는 전체집합 X 에 대응하는 소속 함수 값이 모두 1/2을 만족하는 함수이다. 정의 2.1을 만족하는 퍼지 엔트로피는 일반적인 퍼지 집합 A 에 대한 불확실성을 표현 하는 것으로서 퍼지집합이 일반집합에 근접할 경우, 엔트로피가 0 으로 근접하는 특성을 갖는다. 한 반에 대한 학생들의 점수분포와 중간수준을 일반집합으로 표현하면 그림 1과 같이 나타난다. 그림 1에서 한 반에 대한 점수 분포는 가우시안 특성을 가지며, 이 분포는 중간 성적에 대한 퍼지 소속 함수로 간주할 수 있다. 또한 중간 점수를 취득한 학생에 대한 일반집합을 역시 그림 1에서 같이 나타냈다. 그림 1에 나타낸 결과로서 우리는 가우시안 분포의 밀도함수를 성적이 중간인 학생에 대한 퍼지 소속 함수로 간주하고, 학점 B, C를 취득한 학생에 대하여 일반집합으로 간주한다. 그림에서 확인하듯이 점수가 평균에 가까울수록 소속함수 값이 커지게 되어 소속도가 1에 근접하고 평균에서 멀어질수록 중간수준과는 차이를 나타내므로 소속도는 0으로 접근한다. 또한 퍼지 소속 함수와 일반집합과의 차이 면적을 그림 1에서 진하게 표현 하였는데, 이 면적이 줄어들수록 퍼지 집합은 일반집합에 수렴하게 되어 불확실성이 0로 된다. 따라서 이러한 면적의 차이는 퍼지 집합에 대한 엔트로피로 간주할 수 있다.

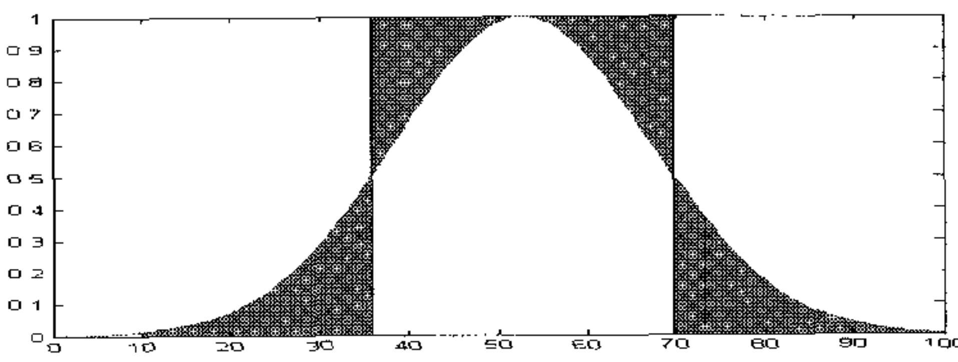

그림 1. 중간수준학생의 소속함수와 B,C 학점의 집합
Fig. 1 Average level student membership function and B, C Grade

퍼지 엔트로피의 정의 2.1을 만족하는 엔트로피는 기존의 결과로서 다음과 같이 제안되어왔다. Fan, Ma 그리고 Xie 의결과를 확인하면, 퍼지 집합 A 에 대한 엔트로피를 다음과 같은 식으로 표현하고 있다[3].

(Fan, Ma, and Xie, 2001) d 를 퍼지 집합 $F(X)$에 대한 σ-거리측도이고 다음의 두가지 성질을 만족하면

(i) $d(\frac{1}{2}D,[0])=d(\frac{1}{2}D,D),\ \forall D\in P(X)$,

(ii) $d(A^{c},B^{c})=d(A,B),\ A,B\in F(X)$,

$e(A)=D(A,A_{near})+1-d(A,A_{far})$ 는 퍼지 엔트로피이다.

(Lee et. al., 2004) d 를 퍼지 집합 $F(X)$에 대한 σ-거리측도이고 다음의 성질을 만족하면
$d(A^{c},B^{c})=d(A,B),\ A,B\in F(X)$,

$$e(A)=2d((A\cap A_{near}),[1])+2d((A\cup A_{near}),[0])-2$$

는 퍼지 엔트로피이다.

(Lee et. al., 2004) d 를 퍼지 집합 $F(X)$에 대한 σ-거리측도이고 다음의 성질을 만족하면
$d(A^{c},B^{c})=d(A,B),\ A,B\in F(X)$,

$$e(A)=2d((A\cap A_{far}),[0])+2d((A\cup A_{far}),[1])$$

는 퍼지 엔트로피이다.

우리는 제안한 퍼지 엔트로피 함수를 이용한 신뢰성 있는 데이터의 추출 시행 결과, 퍼지 엔트로피의 결과만으로는 만족한 결과를 얻지 못하였다. 따라서 보다 신뢰성 있는 결과를 얻기 위하여 통계적 데이터인 평균값을 이용하여 퍼지 엔트로피의 결과를 보완하여 결과를 얻었지만 그 결과 또한 적절한 성능을 내지 못함을 확인하였다 [5]. 이러한 문제점을 극복하기 위하여 새로운 측도의 제안이 요구된다. 따라서 우

리는 다음의 절에서 유사측도를 구성하여 기존의 방법을 보완하고자 한다.

2.2 유사측도

이제 우리는 거리측도를 이용한 유사측도 구성을 위하여 먼저 유사측도에 대한 정의를 소개한다.

정의 2.2 [1] 퍼지 집합 $F(X)$ 또는 일반집합 $P(X)$에 대하여 $s: F^2 \to R^+$ 와 같은 함수가 성립하고 다음과 같은 특징을 만족하면 s는 유사측도라고 정의 한다:

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

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

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

(S4) $\forall A,B,C \in F(X)$ 대하여 $A \subset B \subset C$이면 $s(A,B) \geq s(A,C)$ 이고 $s(B,C) \geq s(A,C)$이다.

정의 2.2에서 유사측도에 대한 특성을 살펴보면, (S1)의 경우 자명하며, 일반집합의 경우 여집합과의 유사도는 0 을 만족한다. 일치한 집합간의 유사도는 최대 유사 값을 나타낸다. 이때 최대 유사 값은 구성하는 유사측도에 의존하며 $R^+ = [0,\infty)$ 를 만족한다. 마지막으로 (S4)에서 우리는 포함관계에 있는 집합간의 유사측도특성을 확인할 수 있다.

기존의 유사측도와 관계된 연구에서는 퍼지 엔트로피, 거리측도와의 관계규명, 그리고 서로간의 특성파악에 집중되어 왔다[1-3]. 퍼지 넘버와 거리측도를 이용한 유사측도의 구성에 대한 연구가 진행하여 유사측도 연산에서의 유용성을 비교하였다 [12]. 이제 유사측도의 정의로부터 우리는 다음의 정리에서 개량된 유사측도를 제안한다. 우리는 그림 1에서 두 집합 사이의 불확실성을 나타내는 엔트로피가 두 집합사이의 거리로 표현되는 면적의 두 배가 됨을 이전의 결과에서 확인할 수 있다 [3,4]. 전체 면적에서 확실성에 비례한 면적을 유사측도로 간주하면 다음과 같이 유사측도를 구성한다.

정리 2.1 어떤 집합 $A,B \in F(X)$ 또는 $P(X)$ 에 대하여 d 가 해밍 거리를 만족하면

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

는 집합 A 와 B에 대하여 유사측도를 만족한다.

증명: 우리는 이제 식 (1) 이 정의 2.2를 만족하는 가를 확인함으로서 증명하려고 한다. (S1) 은 집합 A 와 B의 가환성을 보이는 것인데, 식 (1) 의 구조로부터 자명하다. 그리고 (S2) 를 확인하기 위하여 $\forall D \in P(X)$에 대하여 확인하면

$$s(D,D^c) = 2 - d((D \cap D^c),[1]) - d((D \cup D^c),[0]) = 0$$

이 된다. 왜냐하면 $d((D \cap D^c),[1])$ 와 $d((D \cup D^c),[0])$는 1을 각각 만족한다. 그리고 임의의 집합 A,B 에 대하여 (S3) 는 다음과 같이 증명된다.

$$\begin{aligned} s(A,B) &= 2 - d((A \cap B),[1]) - d((A \cup B),[0]) \\ &\leqq 2 - d((C \cap C),[1]) - d((C \cup C),[0]) \\ &= s(C,C). \end{aligned}$$

여기서 부등식은

$$d((A \cap B),[1]) \geqq d((C \cap C),[1]) \text{ 과 } d((A \cup B),[0]) \geqq d((C \cup C),[0])$$

이 성립하는 것은 자명하다. 마지막으로 (S4) 는 모든 $\forall A,B,C \in F(X)$ 에 대하여 $A \subset B \subset C$ 이면,

$$\begin{aligned} s(A,B) &= 2 - d((A \cap B),[1]) - d((A \cup B),[0]) \\ &= 2 - d(A,[1]) - d(B,[0]) \\ &\geqq 2 - d(A,[1]) - d(C,[0]) \\ &= s(A,C) \end{aligned}$$

이고 또한

$$\begin{aligned} s(B,C) &= 2 - d((B \cap C),[1]) - d((B \cup C),[0]) \\ &= 2 - d(B,[1]) - d(C,[0]) \\ &\geqq 2 - d(A,[1]) - d(C,[0]) \\ &= s(A,C) \end{aligned}$$

가 만족되어 성립한다. 부등식은

$$d(B,[0]) \leqq d(C,[0]) \text{ 와 } d(B,[1]) \leqq d(A,[1])$$

를 통하여 증명된다. 따라서 우리는 식 (1) 이 유사측도를 만족함을 확인할 수 있다. 유사한 방법을 통하여 다음의 정의 역시 유사측도 특성을 만족함을 확인할 수 있다.

정리 2.2 어떤 집합 $A,B \in F(X)$ 또는 $P(X)$ 에 대하여 d 가 해밍 거리를 만족하면

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

는 집합 A 와 B 에 대하여 유사측도를 만족한다.

증명은 정리 2.1의 경우와 유사하게 보일 수 있다.

이상의 정리에서 제안된 유사측도 (1) 과 (2) 는 소속 함수의 형태에 의존하지 않는 유사측도가 됨을 확인 할 수 있다.

3. 유사측도를 이용한 데이터의 추출

제안한 유사측도의 유용성을 확인하기 위하여 우리는 기존에 제시하였던 예제에 동일한 유사측도 연산을 실시하였다 [5]. 3장에서는 먼저 문제의 구성에 대하여 알아보고, 퍼지 엔트로피를 이용한 결과의 문제점을 분석하고 유사측도의 결과를 비교하였다.

3.1 문제의 구성

신뢰성 있는 데이터의 선택문제로서 우리는 다음과 같은 형태의 문제를 제시하였다. 특정 교과목의 수강 후 최종 시험의 결과 다음과 같이 표 1 와 같은 65명 학생의 성적을 얻었다. 통계적 지식을 이용할 경우, 데이터의 평균과 분산 또는 2, 3차 모멘트 등의 데이터를 얻을 수 있다. 통계적 연산의 결과로서 65명의 평균은 52.7 점이고, 표준편차는 14.49 이다. 또한 학점으로 중간수준의 학생을 가늠할 때 B, C학점에 속한점수는 37 부터 71 점까지의 학생이다. 그러나 65명의 학생 중에서 무작위로 추출된 학생이 "어느 정도" 중간수준에 속해있는지는 평균과 분산 등의 통계적 정보만으로는 인지적 지식을 제공하지는 않는다. 따라서 우리는 추출된 학생에 대한 불확실성 또는 확실성에 대하여 계량화하는 것이 요구된다. 연산된 퍼지 엔트로피 또는 유사측도를 이용하여 무작위로 추출한 5명의 학생에 대하여 연산결과를 제시하였다. 그리고 선택된 학생들이 "어느 정도" 중간수준의 학력을

가진 것인지를 비교, 검토하였다. 이를 위하여 우리는 4번의 추출을 시도하였다. 총 4회의 테스트 결과, 우리는 그림 2 ~ 그림5와 같은 결과를 얻었다. 매 테스트에서의 평균값은 52.8, 54.6, 54.6, 그리고 51.4 이다.

표 1. 65명의 점수
Table 1. 65 students' points

students points(65)	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 : 52.70461538, Standard deviation : 14.49286927	

또한 4회의 테스트 결과에 대한 점수와 소속 함수, 그리고 유사측도 값을 표 2에 나타냈다.

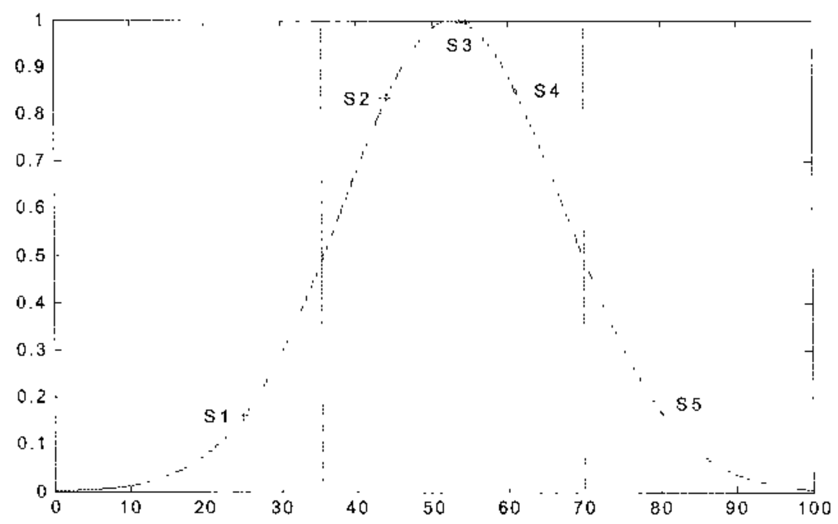

그림 2. 5명 선택(테스트1)
Fig. 2 Selection of 5 students (Test1)

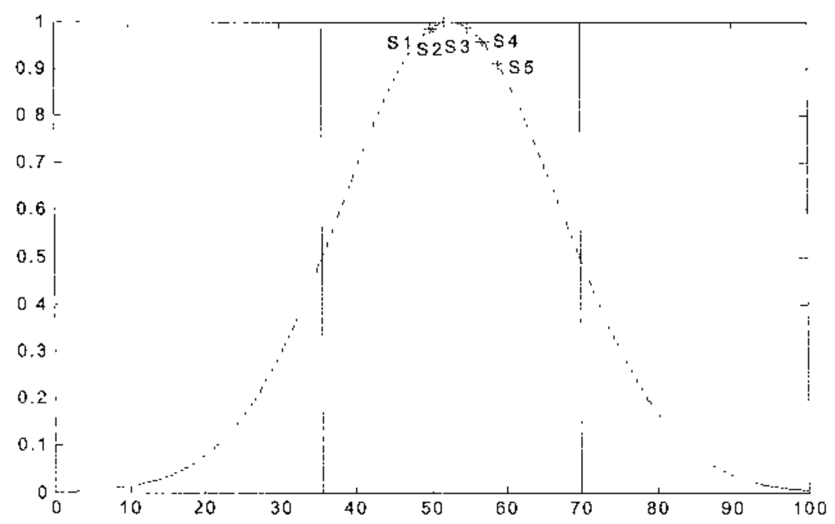

그림 3. 5명 선택(테스트2)
Fig. 3. Selection of 5 students(Test2)

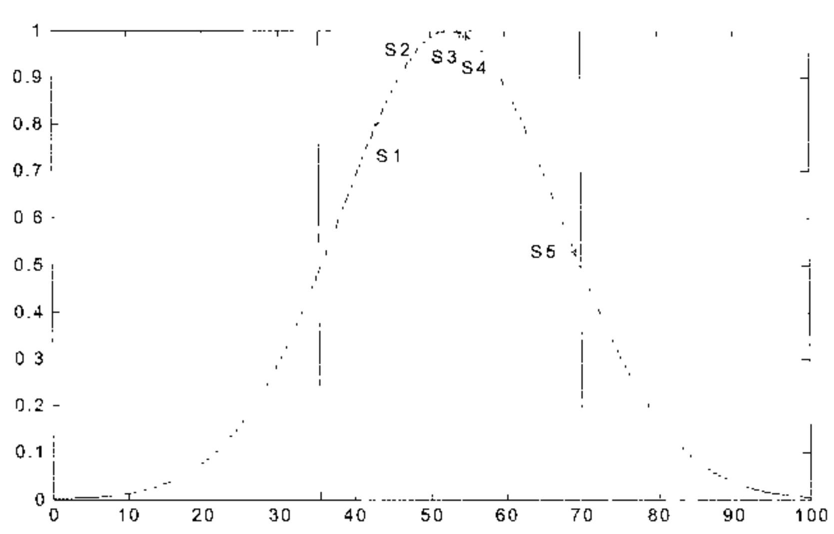

그림 4. 5명 선택(테스트3)
Fig. 4. Selection of 5 students(Test3)

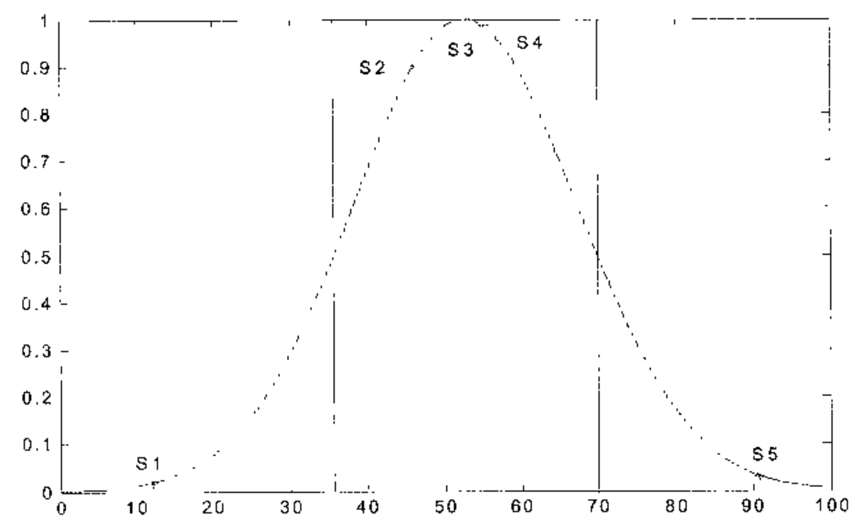

그림 5. 5명 선택(테스트4)
Fig. 5. Selection of 5 students(Test4)

표 2. 점수, 소속 함수 값 그리고 유사측도
Table 2. Sample, Membership and Similarity values

	점수	소속함수값	유사측도
Test 1	25	0.161	0.322
	44	0.835	1.670
	54	0.996	1.992
	61	0.849	1.697
	80	0.170	0.340
평균	52.8	0.60	1.204
Test 2	50	0.983	1.966
	52	0.999	1.998
	55	0.987	1.974
	57	0.957	1.914
	59	0.990	1.980
평균	54.6	0.980	1.966
Test 3	43	0.800	1.600
	52	0.999	1.998
	54	0.996	1.992
	55	0.987	1.974
	69	0.532	1.064
평균	54.6	0.860	1.725
Test 4	12	0.019	0.038
	46	0.899	1.798
	53	1.000	2.000
	55	0.987	1.974
	91	0.031	0.062
평균	51.4	0.590	1.174

3.2 유사측도를 이용한 판단

총 4회의 테스트를 통하여 추출 결과의 유사측도를 계산하였다. 계산 결과 Test 2 의 경우가 가장 높은 유사측도 값을 나타냈다. 그러나 이전의 경우에서는, 퍼지 엔트로피만으로 판단이 용이하지 않았고 통계적 지식을 보완적으로 이용하여 적절한 판단을 할 수 있었다[5]. 참고문헌 5에서 4회 추출의 퍼지 엔트로피는 0.260, 0.066, 0.275 그리고 0.066 이었다. 불확실한 정보에 대한 계량적 수치가 퍼지 엔트로피임을 감안하면 Test 2 와 Test 4 의 추출 결과가 가장 낮은 엔트

로피 값을 나타내고 있다. 두 가지의 추출 중에서 보다 신뢰성 있는 경우를 확정하기 위하여 데이터의 평균 오차를 구한 결과를 표 3에 나타냈다.

표 3. 평균오차 그리고 퍼지 엔트로피
Table3. Average error and Fuzzy entropy

	평균 오차	퍼지엔트로피
Test1	0.1	0.26
Test2	1.9	0.066
Test3	1.9	0.275
Test4	1.3	0.066

데이터의 평균을 고려할 경우, Test 1 의 경우가 가장 평균에 근접하는 결과이다. 그러나 Test 1 의 경우, 퍼지 엔트로피 값이 Test 2 와 Test 3 에 비교하여 현저히 크다. 따라서 두 가지의 측도를 이용하면 가장 신뢰성 있는 선택은 Test 4 의 경우라고 할 수 있다. 그러나 결과를 확인하면, 판정결과는 올바른 판정으로 인정하기가 쉽지 않다. 통계적인 결과인 평균은 주어진 데이터에 의하여 계산되어 지지만, 데이터에 대한 불확실성의 측도인 퍼지 엔트로피의 결과는 신뢰성 있는 데이터의 추출에 크게 기여하지 못함을 확인할 수 있다. 결과에 대한 원인은 그림 3 과 5를 통하여 확인할 수 있다. 그림 5의 경우, s1 과 s5의 경우 평균에서 멀리 떨어진 12, 91점을 획득한 경우로서 두 데이터의 퍼지 엔트로피역시 매우 작음을 확인 할 수 있다. 직접적인 원인은 퍼지 엔트로피의 특성 중에서 (E4)인 $e(A)=e(A^c), \forall A \in F(X)$에 기인한다. 따라서 퍼지 엔트로피만으로 또는 통계적 정보를 부가적으로 이용하더라도 신뢰성 있는 정보의 추출은 쉽게 수행되지 못함을 알 수 있다.

정리 2.1과 2.2를 만족하는 유사 측도를 이용하여 Test 결과에 대한 유사 측도 연산결과를 표2 에 나타냈다. 정규화하지 않은 결과이지만 4번의 추출결과를 비교하면 Test 2 의 추출결과가 가장 높은 유사측도를 나타냄을 알 수 있다. 따라서 Test 2 의 결과가 가장 신뢰성 있는 결과로 나타나며 그림 2 ~그림5를 통하여 결과임을 확인할 수 있다.

4. 결 론

우리는 본 논문에서 신뢰성 있는 데이터의 추출에 대한 연구를 실시하였다. 신뢰성에 대한 계량적 값을 제공하기위한 시도로서 퍼지 엔트로피와 유사측도가 고려되었다. 위를 위하여 퍼지 집합에 대한 해석을 실시하였고, 퍼지 엔트로피에 대한 기존의 결과를 분석하였다. 또한 거리측도를 이용하여 구체적인 형태로 유사측도를 구성하였으며 구성된 유사측도의 적절함을 증명을 통하여 확인 하였다. 제안된 유사측도의 유용성은 기존의 연구와의 비교를 통하여 확인하였다. 얻어진 결과는 결론적으로 퍼지 집합과 일반집합과의 유사도 계산에 활용되었다. 보다 일반적인 유사측도의 구성을 위하여 추후 퍼지집합 간의 유사측도 또는 일반집합 간의 유사측도로 확장하는 연구가 요구된다.

참 고 문 헌

[1] X. Liu, "Entropy, distance measure and similarity measure of fuzzy sets and their relations," *Fuzzy Sets and Systems*, Vol. 52, pp. 305-318, 1992.

[2] J. L. Fan, W. X. Xie, "Distance measure and induced fuzzy entropy," *Fuzzy Set and Systems*, Vol. 104, pp. 305-314, 1999.

[3] J. L. Fan, Y. L. Ma, and W. X. Xie, "On some properties of distance measures," *Fuzzy Set and Systems*, Vol. 117, pp. 355-361, 2001.

[4] S.H. Lee, S.P. Cheon, and Jinho Kim, "Measure of certainty with fuzzy entropy function", *LNAI*, Vol. 4114, pp. 134-139, 2006.

[5] S.H. Lee, Y.T. Kim, S.P. Cheon, and Sungshin Kim, "Reliable Data Selection with Fuzzy Entropy", *LNAI*, Vol. 3613, pp. 203-212, 2005.

[6] S.H. Lee, J.M. Kim, and Y.K. Choi, "Similarity measure construction using fuzzy entropy and distance measure", *LNA*I Vol.4114, pp. 952-958, 2006.

[7] S.M. Chen, "New methods for subjective mental workload assessment and fuzzy risk analysis", *Cybern. Syst. : Int. J.*, Vol. 27, No. 5, pp. 449-472, 1996.

[8] C.H. Hsieh and S.H. Chen, "Similarity of generalized fuzzy numbers with graded mean integration representation," in *Proc. 8th Int. Fuzzy Systems Association World Congr.*, Vol 2, pp. 551-555, 1999.

[9] H. S. Lee, "An optimal aggregation method for fuzzy opinions of group decision," *Proc. 1999 IEEE Int. Conf. Systems, Man, Cybernetics*, Vol. 3, pp. 314-319, 1999.

[10] 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.

[11] P. Subasic and K. Hirota, "Similarity rules and gradual rules for analogical and interpolative reasoning with imprecise data," *Fuzzy Sets and Systems*, Vol. 96, No. 1, pp. 53-75, 1998.

[12] S.H. Lee, "Comparison Study for Similarities based on Distance Measure and Fuzzy Number", *Journal of Fuzzy Logic and Intelligent Systems*, Vol.17, No. 1, pp. 1-6, 2007.

저 자 소 개

류수록(Soorok Ryu)
2003년 : 경북대학교 수학과(이학사)
2005년 : 경북대학교 산업응용수학과(이학석사)
2005년~경북대학교 수학부 박사과정 수료

관심분야 : numerical PDE, optimal control, fluid mechanics
E-mail : sryu@knu.ac.kr

이상혁(Sang-Hyuk Lee)

제 17권 1호 (2007년 2월호) 참조