韓國 OA學會 論文誌
第3卷 第2號, 1998. 6.
98-3-2-2-10

최대 엔트로피 분포를 이용한 퍼지 관측데이터의 분석법에 관한 연구

유 재 휘* 유 동 일**

An Analysis of Fuzzy Survey Data Based on the Maximum Entropy Principle

Jae-Whee Yoo* Dong-Il Yoo**

요 약

통상 통계적인 데이터 해석에서 취급되는 데이터는 확정된 값으로서 통계 처리를 실시한다. 그러나 복잡·대규모화하는 현대의 시스템에 있어서는 정확하게 측정된 데이터만을 취급하는 것은 곤란하며 인간의 주관적인 판단에 따른 데이터를 수집하는 경우가 발생하게 된다. 본 연구에서는 이러한 인간의 주관적인 판단에 따른 데이터를 퍼지 관측 데이터로 하여(언어 변수에 의해 Membership 함수를 정의한다.) 최대 엔트로피 원리를 이용한 새로운 분석 방법을 제안한다. 또한 보다 현실적인 상황 아래 시뮬레이션을 실시함으로서 제안 모델의 유효성을 검증한다.

Abstract

In usual statistical data analysis, we describe statistical data by exact values. However, in modern complex and large-scale systems, it is difficult to treat the systems using only exact data. In this paper, we define these data as fuzzy data(ie. Linguistic variable applied to make the member-ship function.) and propose a new method to get an analysis of fuzzy survey data based on the maximum entropy principle. Also, we propose a new method of discrimination by measuring distance between a distribution of the stable state and estimated distribution of the present state using the Kullback - Leibler information. Furthermore, we investigate the validity of our method by computer simulations under realistic situations.

---

* 여주대학 사무자동화과 조교수
** 여주대학 사무자동화과 조교수

논문접수 : 98.5.10 심사완료 : 98.6.12

1. 서 론

현실 세계에서 인간들이 이용하고 있는 정보에는 불확실성을 내포하는 불완전한 것이 많다. 그런데도 인간은 불확실성을 포함하고 있는 정보를 불확실성을 포함한 채로 취급하기도 하고, 또 불확실한 형태로 사고하고 판단하기도 한다. 이와 같은 사고나 판단 등의 불확실성을 수학적으로 무리 없이 표현하고자 하는 것이 퍼지이론이다. 퍼지이론의 연구는 1965년 Zadeh[1,2]가 퍼지 집합을 제안함으로서 시작되었다. 퍼지 집합은 경계가 불확실한 집합이며, 이것은 인간의 정상적인 사고나 판단 등을 정량적으로 취급하기 위해 제안된 것이다. 퍼지 집합론은 그 후 기초 이론의 발전과 함께 측정, 논리 등으로 확장되고, 더 나아가 모델링, 평가, 최적화, 의사 결정, 제어, 진단, 정보 등의 응용 분야에서의 방법론으로 발전하고 있다. 이와 같이 퍼지 이론이 추구하는 궁극적인 목적은 인간과 같은 정보처리를 실시할 수 있는 시스템의 구축이며, 퍼지이론은 이러한 정서에 맞도록 불확실한 정보를 처리하고 추론하는 것에 적합한 이론이다.

최근에 공학 분야의 응용에 있어서 퍼지이론을 응용하여 언어 데이터를 통계적으로 처리하는 연구가 활발하게 진행되고 있다. 이와 같이 언어 데이터가 가지는 애매함을 취급하기 위하여 언어 변수라는 개념이 제창되었고, 그 처리 방법도 확립되고 있다. 최근의 연구에 있어서는 퍼지 관측에 의해 얻어진 하나의 언어를 라벨을 부여한 퍼지 데이터로 해석하는 것에 그 특징이 있다[3,4]. 여기에서 퍼지 데이터라는 것은 현재 퍼지 관측을 실시한 특성에 대해 상대적인 계량 척도가 숨겨져 있고, Zadeh가 정의한 퍼지 사상의 소산으로서 얻어진 데이터를 가리킨다. 이 경우 임의의 퍼지 데이터를 얻을 확률은 그 퍼지 데이터를 결정하는 Membership 함수와 숨겨진 계량 척도 상에서 구성되는 확률 밀도 함수에 의해 구할 수 있다.

한편, 연속 분포에 대해 해석적인 구조가 미지인 경우 데이터에 의해 분포를 추정할 때, 고전적 접근방법에서는 데이터를 히스토그램화하여 피아송이라 불리는 파라메터의 분포 모델에 대한 적합도를 나타내는 χ^2 값을 추정하는 방법이 있다. 이에 비해 Jaynes[5]는 "편견을 최소화하는 확률 분포는 주어진 정보에 의한 제약 아래 엔트로피를 최대로 하는 분포이다"라고 최대 엔트로피 원리를 명확하게 하였다.

본 연구에서는 이러한 퍼지 데이터 해석의 입장에 서서 앙케이트의 대상이 되는 "평판"이나 "여론"이라는 것에 숨겨진 계량 척도를 상정하고, 그 위에 성립하는 확률 분포를 Jaynes가 제안한 최대 엔트로피 원리를 이용한 퍼지 관측 데이터에 대한 새로운 분석 모델을 제안하고, 관측자를 대상으로 앙케이트를 실시하였을 때 그 대상에 대한 특질에 의해 평가의 분포가 경험적으로 일정한 패턴을 형성하는 경우 퍼지 관측 데이터로부터 측정된 최대 엔트로피 분포와 모델 분포와의 K-L정보량[6,7](Kullback and Leibler Information)을 추정하여 앙케이트 결과 배후의 분포가 어느 모델에 해당되는가를 판별하는 방법에 대해서도 제안한다.

2. 최대 엔트로피 원리에 의한 확률 분포의 추정

본 장에서는 어떤 연속분포 $F(x)$에 대하여 그 해석적인 구조가 미지인 경우, 분포 F에 의한 데이터 $x_1, \cdots, x_n$로부터 $F(x)$를 추정하는 문제에 대한 고전적인 접근방법은 $x_1, \cdots, x_n$를 히스토그램화하여 피아송이라 불리는 다파라메터의 분포 모델에 대한 적합도를 나타내는 χ^2값에 따라 추정하는 방법이 있다. 이것에 대해 $F(x)$가 확률 밀도 함수 $f(x)$를 가질 때, Jaynes는 『편견을 최소로 하는 확률 분포는 주어진 정보에 의한 제약 아래 다음과 같은 엔트로피 를 최대화하는 분포이다』라는 최대 엔트로피 원리를 명확하게 하였다.

$$S = -\int_R f(x) \ln[f(x)]dx \tag{1}$$

한편, Siddall 과 Diab[8,9]는 이 최대 엔트로피 원리에 따라 "주어진 정보"를 데이터로부터 얻을 수 있는 moment로서 해석하고, 데이터에 의한 moment와 F에 의한 moment를 정해진 차수까지 일치 시킨 후 엔트로피를 최대화하는 분포형을 도출하였다. 이것은 표본의 m차까지의 moment가 이미 알려져 있고, 확률 변수의 영역이 정의되어 있을 때, 그 확률 분포를 추정하는 것으로 개략적인 내용은 다음과 같다. 먼저, 주어진 정보는 아래의 제약식에 따른다고 정의한다.

$$\int_R f(x)\,dx = 1 \tag{2}$$

$$\int_R x^j f(x)\,dx = \mu_j, \quad j = 1, 2, \cdots, m \tag{3}$$

여기에서 m은 이용되는 moment 이고, μ_j는 원점에서의 j차 moment이다. 식(1)~식(3)에 관한 문제는 등식 제약인 최대화 문제이므로 Lagrange 승수를 $\lambda_0, \lambda_1, \cdots, \lambda_m$ 이라 하면, 다음의 함수 $\overline{S}$를 최대화하는 문제로 정식화된다.

$$\overline{S} = S + (\lambda_0 + 1)\left[\int_R f(x)dx - 1\right] + \sum_{j=1}^{m} \lambda_j \left[\int_R x^j f(x)dx - \mu_j\right] \tag{4}$$

여기에서 승수($\lambda_0 + 1$)이 λ_0대신 이용되는 것은 보다 편리한 결과를 얻기 위함이다. 식(4)를 $f(x)$에 대해 미분하여 0으로 놓으면 다음과 같은 식을 얻을 수 있다.

$$\frac{d\overline{S}}{df(x)} = -\int_R \{\ln[f(x)+1]\}dx - (\lambda_0 + 1)\int_R dx - \sum_{j=1}^{m} \lambda_j \left(\int_R x^j dx\right) = 0 \tag{5}$$

또, 적분 기호 아래 각 항을 정리하면 식(5)는 다음과 같이 된다.

$$\int_R \left\{-\ln[f(x)] - 1 + \lambda_j + 1 + \sum_{j=1}^{m} \lambda^j x\right\} dx = 0 \tag{6}$$

식(6)이 항등적으로 성립하기 위해서는 { }안이 0이어야 하며, 따라서 식(6)은 다음과 같이 되고, 이것이 최대 엔트로피 분포의 해석적인 형태이다.

$$f(x) = \exp\left(\lambda_0 + \sum_{j=1}^{m} \lambda_j x^j\right) \tag{7}$$

문제는 몇 개의 moment와 x의 범위를 주고, 식(2)나 식(3)에 의해 표현된 $(m+1)$개의 등식을 푸는 것에 의해 λ_j의 값을 결정하는 것이다.

여기서 식(7)을 식(2)에 대입하면 다음 식을 얻는다.

$$\int_R \exp\left(\lambda_0 + \sum_{j=1}^{m} \lambda_j x^j\right) dx = 1 \tag{8}$$

여기에 $e^{-\lambda_0}$를 곱하는 것에 의해 다음과 같이 된다.

$$e^{-\lambda_0} = \int_R \exp\left(\sum_{j=1}^{m} \lambda_j x^j\right) dx \tag{9}$$

따라서 ,

$$\lambda_0 = -\ln \int_R \exp\left(\sum_{j=1}^{m} \lambda_j x^j\right) dx \tag{10}$$

으로 된다.

또, 식(9)를 λ_j에 대해 편미분하면,

$$\frac{\partial \lambda_0}{\partial \lambda_j} = -\int_R x^j \exp\left(\lambda_0 + \sum_{j=1}^{m} \lambda_j x^j\right) dx \tag{11}$$

와 같이 되고, 식(11)과 식(3), 식(7)의 관계로 부터 다음 식을 얻을 수 있다.

$$\frac{\partial \lambda_0}{\partial \lambda_j} = -\mu_j \tag{12}$$

그리고, 식(10)을 λ_j에 대해 편미분하면,

$$\frac{\partial \lambda_0}{\partial \lambda_j} = -\frac{\int_R x^j \exp\left(\sum_{j=1}^{m} \lambda_j x^j\right) dx}{\int_R \exp\left(\sum_{j=1}^{m} \lambda_j x^j\right) dx} \tag{13}$$

이 되고, 식(13)을 식(12)에 대입하여 μ_j로 나누면 다음과 같은 식을 얻는다.

$$1 - \frac{\int_R x^j \exp\left(\sum_{j=1}^{m} \lambda_j x^j\right) dx}{\mu_j \int_R \exp\left(\sum_{j=1}^{m} \lambda_j x^j\right) dx} = 0 \tag{14}$$

그러나, 식(14)를 엄밀하게 만족시키는 해를 얻는 것은 용이하지 않다. 여기서 식(14)의 우변을 R_j로 두고,

$$R_j = \sum_{j=1}^{m} R_j^2 \rightarrow \min z9 \tag{15}$$

으로 하는 것에 의해 비선형 계획법의 적용이 가능한 형식으로 해를 구할 수 있게 된다. 해는 $R \leq \varepsilon$ 또는 모든 j에 대해 $R_j \leq |\varepsilon|$일 때 구할 수 있다. 여기서 ε은 정해진 허용 오차이다. λ_0는 식(10)에 의해 얻을 수 있다. 단, 추정하는 확률 분포 함수의 경계에 대해서는 이미 알려져 있는 것으로 가정된다.

3. 퍼지 관측 데이터와 통계적 처리

3-1. 퍼지 관측 데이터와 Membership 함수

앙케이트에 의한 여론 등 조사의 대부분은 언어 형식의 항목을 선택하게 된다. 즉, 인간의 주관을 언어 값의 형식으로 표현하는 것을 언어 데이터라고 한다. 언어 데이터는 그 값이 수치가 아니라 문장 또는 구로서 표기된다는 점에서 퍼지 데이터와 다르다. 그러나, 언어 데이터 배후에 있는 척도에 따른 확률 분포가 존재하고, 인간의 주관을 표현한 Membership 함수에 따라 언어 데이터가 출현한다는 입장에서 보면 언어 데이터는 라벨 부여 퍼지 데이터라 할 수 있다. 통상, 관측치 또는 실현치는 어느 한점 x_i로서 취급하지만, 본 논문에서는 구간 $H_i = [x_i - h/2, x_i + h/2]$에서 정의된 퍼지 사상(퍼지 집합) 을 퍼지 관측 데이터로 한다. 이러한 퍼지 사상 X_i의 Membership 함수를 x_i로 하고, $\chi_i(x_i) = 1$로 한다. 여기에서 퍼지 관측에 의해 얻어진 퍼지 사상 X_i를 퍼지 관측치하며, 이것을 퍼지 데이터로 한다.

한편, 퍼지 사상과 Membership 함수는 1대1로 대응되므로 이후 X_i와 x_i를 동일시하고, 표시의 복잡성을 피하기 위해 통일적으로 x_i를 이용한다.

한편, $f(x)$를 모집단의 확률 밀도 함수라 하면, 퍼지 데이터 x_i가 출현할 확률은 Zadeh의 정의에 의해 다음과 같이 정의할 수 있다.

$$P_i = \int_{x_i - h/2}^{x_i + h/2} \chi_i(x) f(x)\, dx \tag{16}$$

다음에 퍼지 관측의 기본형으로서, 구간 $[-h/2, h/2]$상에서 정의된 Membership 함수로서 다음과 같은 삼각형을 생각할 수 있다.

$$\chi(x) = -\frac{2}{h}|x| + 1 \tag{17}$$

여기서 $\chi_i(x)$는 $x = 0$을 중심으로 양측에 $h/2$의 폭으로 애매함이 존재하고 있는 것을 나타낸다. 즉, $\chi_i(x)$는 퍼지 관측에 따라 값 x가 발생하였다고 생각되는 주관적인 신뢰도를 나타내고 있다.

한편, 퍼지 데이터를 통상의 통계적인 계산 방법으로 데이터를 처리 하고자 할 때 퍼지 집합을 스칼라화 한다. 즉, Membership 함수의 대표치를 결정할 필요가 있다. Membership 함수 $\chi_i(x)$의 대표치 x_i를 정하는 방법 여러 가지 있지만, 본 연구에서는 Membership 함수 $\chi_i(x)$가 1과 같은 점을 기본 변수의 값으로 취하는 모드법(즉, $x_i = \{x \mid x_i(x) = 1\}$) [10]을 사용한다.

3-2 퍼지 관측데이터의 moment수정

추정 문제를 취급하는 경우, 대표치를 이용하여 처리하면 편차가 발생하므로 이 편차를 제거하기 위해 수정을 실시할 필요가 있다. 이것에 대해서는 奧田[11]등이 다음과 같은 수정 식을 제안하고 있다. 여기에서 μ_i'은 수정 후원점에서의 moment, $\overline{\mu_i'}$는 수정 전 원점에서의 moment를 나타낸다.

$$\mu_1' = \overline{\mu_1'} \tag{18}$$

$$\mu_2' = \overline{\mu_2'} - \frac{h^2}{24} \tag{19}$$

$$\mu'_3 = \overline{\mu'_3} - \frac{h^2}{8}\overline{\mu'_1} \tag{20}$$

$$\mu'_4 = \overline{\mu'_4} - \frac{h^2}{4}\overline{\mu'_4} + \frac{h^4}{160} \tag{21}$$

그러나, 최대 엔트로피 분석에 의한 확률 분포의 추정 알고리즘에서는 평균치에서의 moment를 이용하므로 평균치에서의 moment에 대한 수정 식을 도출할 필요가 있다.

먼저, 원점에서의 moment로부터 평균치로의 moment로의 변환은 다음과 같이 나타낼 수 있다.

$$\mu_1 = \mu'_1 \tag{22}$$

$$\mu_2 = \mu'_2 - \mu_2'^{2} \tag{23}$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu_1'^{3} \tag{24}$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu_1'^{2} - 3\mu_1'^{4} \tag{25}$$

여기에서 μ_i는 평균치에서의 moment이고, μ'_i는 원점에서의 moment를 나타낸다.

역으로 평균치 주변의 moment로 부터 원점 주변의 moment로의 변환은 다음과 같이 된다.

$$\mu'_1 = \mu_1 \tag{26}$$

$$\mu'_2 = \mu_2 + \mu_1^2 \tag{27}$$

$$\mu'_3 = \mu_3 + 3\mu_2\mu_1 + \mu_1^3 \tag{28}$$

$$\mu'_4 = \mu_4 + 4\mu_3\mu_1 + 6\mu_2\mu_1^2 + \mu_1^4 \tag{29}$$

여기에서 식(18)~식(21)를 식(22)~식(25)에 대입하고, 식(26)~식(29)를 이용하면 다음과 같은 수정식을 얻을 수 있다.

$$\mu_1 = \overline{\mu_1} \tag{30}$$

$$\mu_2 = \overline{\mu_2} - \frac{h^2}{24} \tag{31}$$

$$\mu_3 = \overline{\mu_3} \tag{32}$$

$$\mu_4 = \overline{\mu_4} - \frac{h^2}{4}\overline{\mu_4} + \frac{h^4}{160} \tag{33}$$

3-3 여론의 배경에 존재하는 확률분포의 추정 알고리즘

앙케이트 형식에 의한 넓은 의미의 여론 조사에서는 어떤 한 항목에 대한 의견을 수렴할 때, 찬반의 여부 또는 만족 불만족이라는 형태의 질문은 별로 없으며, 통상 부사구를 병용하는 것에 의해 선택할 수 있는 폭을 증가시키고, 미묘한 의견의 차이를 도출하고자 하는 일이 많다.

이러한 부사가 나타내는 정도의 차나 느낌이라는 것은 그 언어를 커뮤니케이션 수단으로서 공유하고 있는 국민, 민족간에서는 이른바 공통의 인식으로서 육성되고 있다고 생각해야 한다. 따라서, 이러한 부사에 수식된 말은 애매하지만 그것은 그 언어를 사용하는 집단에 의해서 공통적인 애매함이 형성되고 있다고 생각한다면, 앙케이트 항목의 선택 항목이 가지는 단어의 언어 상 애매함은 그 집단에 공통적인 퍼지 집합으로 표현할 수 있게 된다.

예로서, 어떤 사건에 관한 성취도에 대한 회답을 선택 항목으로서 5개를 가정하고, 그 Membership 함수를 그림 1과 같이 정의할 수 있다.

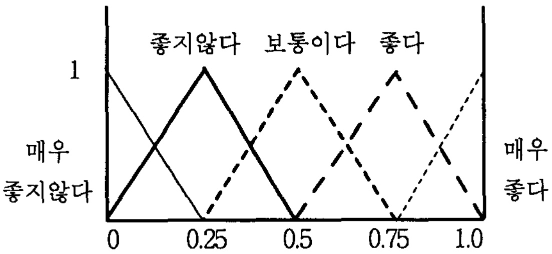

그림 1. 각 항목의 Membership 함수

본 연구에서는 이들의 앙케이트 조사 결과와 각 항목의 Membership함수를 이용하여 배후에 존재하는 확률 밀도 함수를 추정하고, 그 확률 밀도 함수를 이용한 앙케이트 조사 분석을 실시하는 것을 목적으로 하고 있다.

이러한 확률 밀도 함수는 최대 엔트로피 분포를 이용하여 구할 수 있다. 그 알고리즘을 요약하면 다음과 같다.

[순서1] 각 선택 항목에 할당된 Membership 함수를 모드법을 이용하여 대표치화하고, 그 수치에 따라 적당한 차수까지 원점에서의 moment를 계산한다.

[순서2] 계산된 각 moment에 대해 수정을 실시한다.

[순서3] 수정된 각 moment에 대해 최대 엔트로피 분포를 구한다

3-4 앙케이트 배후에 있는 확률 분포의 판별법

본 절에서는 관측자에 대한 평가를 앙케이트에 의해 수집했을 때, 그 대상에 대한 특질에 의해 평가의 분포가 경험적으로 일정한 패턴을 형성하는 경우, 당해 대상이 어느 패턴에 속하고 있는가를 판별하는 문제를 다뤄보기로 한다. 일반적으로 두 개의 분포 $f(x)$, $g(x)$와의 거리를 측정하는 개관적인 척도로서 다음과 같이 정의하고 있다.

$$I(g;f) = \int_{-\infty}^{\infty} g(x)\ \ln\left(\frac{g(x)}{f(x)}\right)dx \qquad (34)$$

식(34)는 $f(x)$에 대한 $g(x)$의 K-L정보량으로 불린다. K-L정보량은 항상 프러스값 또는 0의 값을 가지며, $f(x) \equiv g(x)$일 때, $I(g;f)=0$이 된다. 여기서 $I(g;f)$의 값이 0 에 가까울수록 $f(x)$는 $g(x)$에 가까운 분포라고 할 수 있다. 식(34)는

$$I(g;f) = \int_R \ln\ g(x)\,dG - \int_R \ln f(x)\,dG \qquad (35)$$

로 변형되므로 $g(x)$에 최대 엔트로피 분포식 고려하면 식(35)의 우변 첫 항은 다음과 같이 되고, 계산이 용이하게 된다.

$$\int_R \ln\ g(x)\,dG = \int_R\left(\sum_{j=0}^{m}\lambda_j\ x^j\right)dG$$

$$= \sum_{j=0}^{m}\lambda_j \int_R x^j dG = \sum_{j=0}^{m}\lambda_j\ \mu_j \qquad (36)$$

$f(x)$도 최대엔트로피 분포 형식으로 표현하면 같은 방법으로 계산이 용이하게 되나 본 연구에서는 관측자들에 대한 평가의 분포가 경험적으로 일정한 패턴을 형성하고 있다고 가정하고

$f(x)$를 그림 2에 나타낸 호평형, 불평형, 평범형, 양극형, 대호평형, 대불평형의 여섯 패턴의 모델 분포로 가정한다.

이들 여섯 분포 모델에 대해 앙케이트 조사로부터 다음식에 따라 모델 M_i로 판별하게 된다.

$$\min\ I(g;\ M_i) \qquad i=1,\ldots\ldots,\ldots 6 \qquad (37)$$

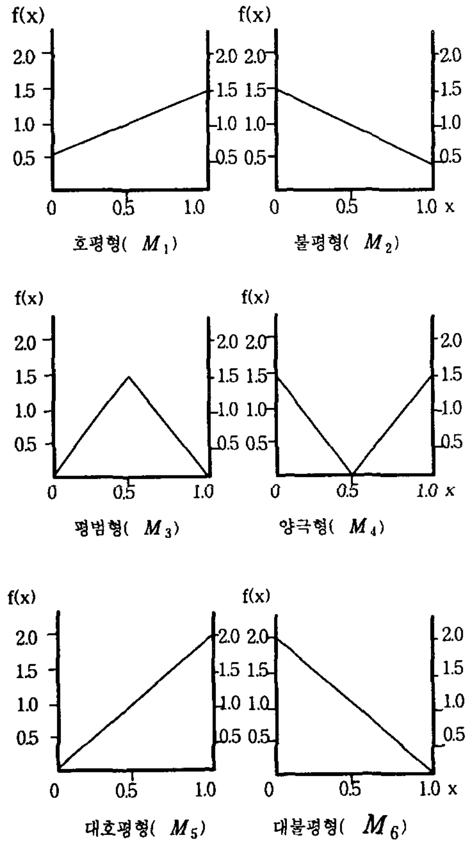

그림 2. 분포모델의 형상(M_1 ~ M_6)

4. 최대 엔트로피 분포에 의한 앙케이트 조사의 분석 예

본 절에서는 표 1에 있는 강의 평가 설문지를 작성하여 학생들로부터 수집된 앙케이트 조사 결과를 제안한 방법을 이용하여 분석함으로서 제안 모델의 유효성을 검증한다. 100명의 학생들을 대상으로 강의

표1. 강의 이해력 평가 설문지 예

본 설문지는 수강한 과목에 대하여 여러분의 솔직한 견해를 수렴하기 위해 만들어 졌습니다. 여러분의 견해는 담당 교수님이 다음 학기 수업을 준비하시는 데 귀중한 자료로 활용될 것 입니다. 해당 번호에 V표 하여 주십시오.

◉ 과목명 : ◉담당교수

◉ 수강학생 : 학과 학년 반 ◉작성일

◉ 평가척도 ① 매우 좋다 ② 좋다 ③ 보통이다 ④ 좋지않다 ⑤ 매우 좋지않다

1. 강의 목표와 방향이 비교적 분명했다.	①	②	③	④	⑤
2. 강의 진행은 강의 계획서의 틀에서 크게 벗어나지 못했다.	①	②	③	④	⑤
3. 학생들에게 요구되는 강의 준비 등 학습 부담의 수준은 적절했다.	①	②	③	④	⑤
4. 교수님은 강의를 충실하고 짜임새 있게 운영하셨다.	①	②	③	④	⑤
5. 교수님은 과제 및 발표 시험 결과 등에 적절한 평가를 하셨다.	①	②	③	④	⑤

평가에 대한 앙케이트 조사를 실시하여 표 2와 같은 결과를 얻을 수 있었다.

표 2의 앙케이트 결과를 이용하여 각 항목에 할당된 Membership 함수를 모드법을 이용하여 대표치화 하고 원점에서의 Moment를 산출한 후, 3-2의 논의에 따라 Moment 수정을 실시하였다.

다음에 2.의 논의에 따라 응답자 배후에 존재하는 확률분포를 추정한 후 그림 2의 분포모델 형상과 K-L 정보량을 측정하였다. 측정 결과를 표 3에 나타내었다. 표 3에서 보면 설문1의 "강의 목표와 방향이 비교적 분명했다"에 대해 응답자 배후에 존재하는 분포는 호평형의 분포에 가장 근접하고 있으며, 같은 방법으로 설문2는 불평형으로 설문 3과 설문 5는 호평형으로 설문 4는 양극형으로 판별됨을 알수 있다.

5. 결 론

본 연구는 앙케이트 등으로 관찰된 언어 데이터를 퍼지 관측 데이터로 보고 그 대표치에 의한 데이터에 Moment 보정을 실시하는 것에 의해 최대 엔트로피 분포를 이용할 수 있게 되고, 인간이 지니고 있는 감각의 배후에 있는 분포를 추정하는 방법을 제안하였다. 그리고, 조사 경향에 따른 모델 분포를 상정하고, 그 분포와 제안 모델에 의해 추정된 분포와의 K-L 정보량을 구하여 앙케이트 결과 추정된 배후의 분포가 어느 모델 분포에 해당하는가를 판별하는 방법을 제안하고 수치 검증을 통해 제안 모델의 유효성을 검증하였다. 한편, 서로 다른 시간 공간에서 앙케이트를 채취하였을 경우 각각의 Membership 함수가 정확하게 정의되어 있지 않으면 제안법을 이용한 분석 결과 자체의 신뢰성이 떨어질 수도 있다. 이것에 대한 좀 더 깊은 연구가 요구된다.

표2 앙케이트 결과 집계표

	매우 좋지않다	좋지 않다	보통 이다	좋다	매우 좋다
설문1	7	15	36	32	10
설문2	16	17	37	23	7
설문3	5	9	50	22	14
설문4	15	6	19	26	34
설문5	15	9	24	25	27

표3. 최대 엔트로피 분포와 모델 분포 (M1 ~ M6)와의 K-L 정보량

모델 / 설문	호평형	불평형	평범형	양극형	대호평형	대불평형
설문 1	0.067816*	0.187843	0.144904	0.643678	0.170897	0.435960
설문 2	0.149148	0.074664*	0.285595	0.570684	0.505024	0.201105
설문 3	0.237988*	0.398983	0.323483	1.112700	0.293786	0.727529
설문 4	0.147820	0.528931	1.181730	0.146769*	0.406872	1.257550
설문 5	0.101912*	0.319611	0.850828	0.161749	0.327551	0.897077

References

[1] Zadeh, L. A. : "Probability Measure of Fuzzy Events" , J. of Math. Appl. Vol. 28, pp.421~427, 1968.

[2] Zadeh, L. A. :"The Concept of a Linguistic Variable and its Apprication to Approximate Reasoning (I) : (Ⅱ)", Inform. Sci. 8, pp. 199~249 & 9, pp. 43~80, 1975.

[3] Kanagawa, A., Tamaki, F. and Ohta, H. : "Fuzzy Control Charts for Linguistic Data", Fuzzy Engineering toward Human Friendly Systerms, Vol. 2 proc. of IFES '91 (YOKOHAMA), pp.644~654, 1991.

[4] Kaufmann, A. and Gupta, M. M. : Introduction to Fuzzy Arithmetic, Van Norstrand Reinhold, 1985.

[5] Jaynes, E.T. : "Information Theory and Statistical Mechanics", phys. Rev., Vol.106, pp. 620~630, 1957.

[6] Kullback, S. : "An Application of Information Theory to Multivariate Analisys Ⅱ", Ann. Math. Statist., Vol.27, pp. 122~146, 1956.

[7] Kullback, S. and Leiber, R. A. : "On Information and Sufficiency", Ann. Math. Statist., Vol. 22, pp. 76~86, 1951.

[8] Siddall, J. N. and Diab, Y. : "The Use Probabilistic Design of probability Curves Generated by Maximizing the Shannon Entropy Function Constained by Moments", Transactions of SME Ser.B, Vol. 97, No. 3, pp. 843~852, 1975.

[9] Siddall. J. N. : Probabilistic Engineering Design -Principles and Application, Marcel Dekker, Inc, 1983.

[10] Kupperman, M. :"Further Applications of Information Theory to Multivariate Analysis and Statistical Inference", Ann. Math. Statist., Vol. 27, pp. 1172~1184, 1958.

[11] 奥田徹示, 古殿幸雄, 淺居喜代治:"フアジィ 観測データに基づく統計的推定 ", 計測自動制御學會論文集, Vol. 26, No.5, pp. 564~571, 1990.

● 저자소개

유 재 휘

1982년 ~ 1986년 : 명지대학교 전자계산학과 (공학사)
1986년 ~ 1988년 : 명지대학교 대학원 전자계산학과 (공학석사)
1992년 ~ 현재 : 명지대학교 대학원 컴퓨터공학과 박사과정 수료
1995년 ~ 현재 : 여주대학 사무자동화과 조교수

유 동 일

1985년 ~ 1989년 : 한남대학교 상업교육과 (공학사)
1991년 ~ 1993년 : 일본 오사카부립대학 대학원 경영공학 (공학석사)
1993년 ~ 1995년 : 일본 오사카부립대학 대학원 경영공학 (공학박사)
1995년 ~ 현재 : 여주대학 사무자동화과 조교수