

퍼지집합모델

퍼지 논리는 자연 언어 등의 애매함을 정량적으로 표현하기 위하여 1965년 미국 버클리대학교의 L. A. 자데(Zadeh) 교수에 의해 도입된 퍼지 집합의 사고방식을 기초로 하고 있다.

퍼지 집합의 개념은 각 대상이 어떤 모임에 속한다 또는 속하지 않는다는 이진법 논리로부터 벗어나, 각 대상이 그 모임에 속하는 정도를 **소속함수(membership function)**로 나타내고 그 소속함수를 대응되는 대상과 함께 표기하는 집합이다.

'만약 토마토가 붉다면 그 토마토는 잘 익은 것이다' '이 토마토는 매우 붉다. 따라서 이 토마토는 매우 잘 익었다' 라는 예를 들어보자. 이것은 우리들이 일상에서 보통 행하고 있는 추론이다. 이와 같은 추론을 종래의 **BIM 논리**에 기초한 기호논리학으로 설명하는 것은 곤란하다. 그 이유는 전제 자체가 애매하며 동시에 붉다거나 매우 붉다고 하는 말은 퍼지한 의미로 쓰여지고 있기 때문이다.

+ 문헌과 쿼리의 matching은 근접 또는 **애매한 matching**으로 이루어진다.

+ 퍼지 집합 A에 x의 소속 정도 $\mu_A(x)$:

- $\mu_A(x) : X \rightarrow [0,1]$
 , X : 전체 집합(universal set)
 , [0,1] : 0과 1 사이의 실수

+ Fuzzy logic operators

전통적인 집합 개념과 같이, 퍼지 집합에서도 여집합, 합집합 등의 집합의 연산을 정의할 수 있다.

- + 퍼지 여집합: $\mu_{A^c}(x) = 1 - \mu_A(x)$
- + 퍼지 합집합: $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$
- + 퍼지 교집합: $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$
- + 퍼지 차집합: $A - B = A \cap B^c$

+ 퍼지 데이터베이스 검색 예

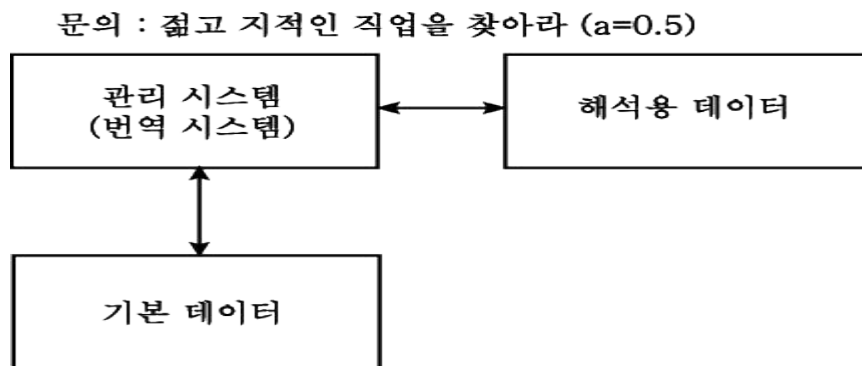
데이터베이스에 퍼지이론의 사고방법을 도입해 보자.

성명과 연령, 직업 일람표가 있다고 치자(그림 1). 지금까지의 방법대로라면 성명을 넣으면 연령과 직업이, 또는 연령이나 직업을 넣으면 성명 일람표가 출력된다 이와 같은 방법에서 '젊고 지적인 직업인'을 찾고 싶다고 하자. 지적이란 것과 젊다는 것은 언어로서 데이터베이스에는 나타나지 않음으로 키워드로써 사용할 수가 없다. 지적이니 젊다느니 하는 말은 주관에 따라 달라지는 **애매한 정보**이다.

성명	연령	직업
A1	15	학 생
A2	60	사 장
A3	20	세일즈맨
A4	28	기 술 자
A5	30	교 사

<그림 1> 기본 데이터

이와 같은 예를 <그림2>에서 퍼지이론의 생각하는 방법을 도입, 퍼지데이터베이스의 구성예를 소개한다. <그림 2>에서 기본 데이터라고 표시되어 있는 것이 주어진 데이터의 일람표(그림1)를 나타내고 있다. 통상 데이터베이스라도 문의를 해석하기 위한 관리 시스템이 있는 반면에 퍼지데이터베이스는 해석용 데이터라는 것이 새롭게 추가된다.



<그림 2> 퍼지데이터베이스의 예

관리시스템에서는 '젊고 지적인 직업인을 찾아라' 라고 하는 요구가 있을 때 젊다고 하는 언어가 연령이라고 하는 집합상의 퍼지집합이며 지적인 직업이라고 하는 말이 직업이라고 하는 집합상의 퍼지집합이라는 것을 알 수 있다(실제 요구는 연령=젊다, 직업=지적직업 등과 같이 적는다). 다음에 젊다라든지 지적직업을 해석용 데이터 중에서 찾는다.

여기에는 각각의 퍼지집합의 멤버집합수가 <그림 3>처럼 기억되어 있다. 연령과 같이 연속적으로 소속척도가 바뀌는 멤버집합수에서는 일반적으로 <그림 3>(a) 처럼 대표점에 대한 수치만을 표시해 두고, 도중의 점에 대한 값은 線形補間이라 하여, 그 점을 포함하는 양끝의 점을 수치에서 비례계산하는 것이 보통이다.

연령	μ	직업	μ
0	1	사장	0.6
20	1	세일즈맨	0.3
25	0.8	기술자	0.8
30	0.5	교사	0.9
35	0	학생	0.5

(a) 젊다

(b) 지적직업

<그림 3> 번역용 데이터멤버십관수

기본데이터의 각 사람에 대해 젊고 지적인 직업의 척도를 계산해 보자.

A 씨는 15세 이니까 젊다고 하는 멤버십함수의 값은 <그림 3>(a)에 의해 1이 된다. 학생이기 때문에 지적인 직업이라고 하는 면에서는 <그림 3>(b)에서 0.5가 된다. 젊고 지적인 직업이란 젊고 동시에 지적인 직업이란 의미이다. '동시에'에 해당하는 연산은 단순히 작은 쪽을 취한다는 것이 되기 때문에 A1 씨의 이 요구에 대한 일치도는 1과 0.5의 작은 쪽이니깐 0.5가 된다. A2씨에 대해서는 젊다고 하는 소속도가 0이기 때문에 결과는 곧 0이다.

마찬가지로 A3, A4 및 A5 씨는 각각 0.3, 0.62 및 0.5가 된다.

A4	28	기술자	0.62
A5	30	교사	0.5
A1	15	학생	0.5

<그림 4> 문의에 대한 회답

A4씨가 0.62가 되는 것은 연령 28세의 소속도를 <그림 1>(a)에서 線形補間에서 구한 값이다.

그럼 이제 출력을 보자. 성명, 연령 및 직업과 함께 출력해 보면, <그림 2>의 요구에서 a=0.5라고 표기되어 있는 것은 일치도가 0.5이상의 것을 출력하라는 것이다. 이 값은 자유로이 설정할 수 있다 그러므로 이 경우의 답은 <그림 4>과 같이 일치도의 크기순에 따라 세 가지로 출력되어 A4가 0.62로 가장 일치도가 높은 것을 알 수 있다.

이상, 극히 간단한 예로 퍼지데이터베이스의 사고방법을 소개했다.

그러면 해석용 데이터 정의를 누가 내리는가를 생각해 보자. 물론 사용자가 개인의 주관에 의해 임의로 정해도 되는 것이다. 일반적으로는 표준인 해석용 데이터는 준비해 두고 사용하는 사람에 따라 마음에 들지 않으면 자유로이 바뀌어도 된다. 또한 해석용 데이터에 쓰고 싶은 언어에 대한 퍼지집합이 입력되어 있지 않을 때는 추가설정이 가능하도록 해 둘 수 있다.

<끝>