퍼지 라프 관계 모델에 관한 연구

A Study on Fuzzy Rough Relational Model

정 홍, 김정호

Hong Chung, Jung Ho Kim

계명대학교 컴퓨터.전자공학부

Department of Computer Engineering, Keimyung University

Abstract

The conventional relational databases have difficulties to efficiently represent various kinds of data because an attribute of a tuple should have only one elementary value. In order to represent ambiguous and imprecious information, fuzzy set and rough set have been gaining acceptance, especially as a tool for knowledge discovery in databases. One of former researches applies only one fuzzy membership value to a tuple. We suggest a more advanced model for data representation by way of applying many membership values to a tuple, i.e. one membership value to each attribute of a tuple.

1. 서론

데이터베이스에서 정확한 값의 데이터 처리는 물론, 부정확하고 모호한 데이터를 처리할 필요성이 점점 증대되고 있는데, 이를 위해 퍼지집합(fuzzy set)이 모호함을 표현하는 유용한 도구로 이용되고 있다. 이는 관계 데이터베이스와 결합되어 속성값과 질의의 모호성간에 관련성을 표현하는 우수한 메카니즘으로 인정되고 있다[2]. 그리고 최근 라프집합(rough set) 이론[5]이 데이터베이스에서 지식 발견을 위한 도구로 각광을 받고 있는데, 이는 데이터에서 패턴(지식)을 발견하기 위해 모호성이나 불확실성을 제어하는 유한집합, 동치관계 등의 수학적 이론을 바탕으로 한다[4].

지금까지 사용되어 오고 있는 관계 데이터베이스는 하나의 속성에 하나의 값만을 가질수 있는 정형화된 형태로, 다양한 정보들을 효율적으로 표현하기가 힘들다[8,9]. 기존의 퍼지라프 관계모델[1]은 투플에 하나의 소속도를 두어 데이터의 소속정도를 알아내는 것이었으나 본 논문에서는 각각의 속성값 마다 소속도를 부여하여 한 단계 진일보한 형태의 데이터 표현 방법을 제시하고자 한다.

2. 관련연구

2.1 퍼지 집합

퍼지집합은 1965년 Zadeh[7]에 의하여 제안되었고, Mamdani[3]에 의해 처음 제어에 이용되었다. 전체집합 X가 있고 퍼지집합 A가 주어지면, 이 함수는 값 $\mu_A(x)$를 모든 $x \in X$에 다음과 같이 할당한다.

$$\mu_A: X \rightarrow [0,1]$$

여기서 [0,1]은 실수 0에서 1까지의 범위를 나타낸다.

A와 B가 소속함수 μ_A, μ_B를 가진 퍼지집합이라 할 때, 퍼지 합집합과 퍼지 교집합은 다음과 같이 정의한다[7].

$$\mu_{A \cup B} = \mathrm{MAX}(\mu_A, \mu_B)$$

$$\mu_{A \cap B} = \mathrm{MIN}(\mu_A, \mu_B)$$

2.2. 라프집합

라프집합은 1982년 Pawlak[5]에 의해 제안되었는데, 이는 부정확하고 불완전한 정보를 분류하는 문제와 데이터베이스에서 원인, 결과를 데이터베이스 학습 형태로 인식하고자 하는 것이다[1].

U를 전체집합, R을 U에 있는 동치관계라 할 때, A=(U,R)를 근사공간이라 한다. $x,y \in U$, $(x,y) \in R$일 때 x와 y를 A에서 불분간이라고 한다. X를 U의 부분집합이라 할 때 A에서 X를 최소한 포함한 집합을 A에서 X의 상한근사라 하며 $R_U X$로 표기하고, A에서 X를 최대한 포함한 집합을 A에서 X의 하한근사라 하며 $R_L X$로 표기하며, 다음과 같이 정의한다.

A에서 X의 하한근사 : $R_L X = \{x \in U \mid [x]_R \subseteq X\}$

A에서 X의 상한근사 : $R_U X = \{x \in U \mid [x]_R \cap X \neq \emptyset\}$

여기서 $[x]_R$은 U의 원소 x에 대해 X를 포함하는 R의 동치 클래스이다.

그리고 R_LX를 R에 대한 X의 양영역, U - R_UX를 음영역, R_UX - R_LX를 경계 영역이라 한다. $R_LX=R_UX$ 이면 R-definable이라 하며, $R_LX \neq R_UX$ 이면 X는 R에 대해 라프라 한다.

2.3. 퍼지라프집합

라프집합은 음영역, 경계영역, 양영역을 표현하기 위해 퍼지 소속함수 μ를 {0, 0.5, 1}로 나타낼 수 있다[6]. 즉, 하한근사 혹은 양영역의 모든 원소는 1의 소속도를 갖는다. 하한근사 영역에 포함되어 있지 않은 상한영역 즉, 경계영역의 모든 원소는 0.5의 소속도를 가진다. 라프집합에 포함되지 않은 원소는 0의 소속도를 가진다.

퍼지성을 라프 관계 모델에 결합시키는 목적은 퍼지 소속도를 사용하여 경계영역의 라프성 정도를 수량화하고자 하는 것으로 경계영역의 요소들의 소속도가 0.5로 고정되지 않고 0과 1 사이의 값(0, 1은제외)을 갖도록 하는 것이다[1].

U가 전체집합이고 X가 U의 부분집합일때, U의 퍼지라프집합 Y는 소속도가 다음과 같은 소속함수 $\mu_Y(x)$ 이다.

$$\mu_Y(R_LX) = 1$$
$$\mu_Y(U-R_UX) = 0$$
$$0<\mu_Y(R_UX-R_LX)<1$$

즉, 양영역의 모든 요소는 1의 소속도를 가지고, 경계영역의 요소는 0과 1 사이의 소속도를 가진다.

퍼지라프집합의 합집합과 교집합 연산자는 보통의 퍼지집합처럼 중복 요소의 소속도를 구하기 위해 MIN, MAX를 사용한다.

3. 퍼지라프 관계모델

퍼지라프 관계모델은 일반 관계모델처럼 튜플을 가진 관계들의 집합으로 데이터를 표현한다. 퍼지라프 관계 R은 $(2^{D1},\mu_1)\times(2^{D2},\mu_2)\times...\times(2^{Dn},\mu_n)$의 부분집합이다. 튜플 t_i는 $(d_{i1},\mu_{i1}),(d_{i2},\mu_{i2}),...,(d_{in},\mu_{in})$의 형태를 가지는데, d_{ij}는 특정 정의역 집합 D_j에서의 정의역 값이고, 퍼지 소속도는 $\mu_{ij}\in\mu_j$이다.

예를들어 D_1이 색깔, D_2가 모양, D_3이 재질인 다음과 같은 퍼지라프 관계가 있다고 하자.

객체	색 깔	모 양	재 질
A	{(빨강,.5),(노랑,.2),(파랑,.2),(초록,.1}	{(원형,.5),(사각,.3),(삼각,.2)}	{(나무,.3),(플라스틱,.7)}
B	{(빨강,.5),(노랑,.3),(파랑,.2)}	{(원형,1)}	{(나무,.7),(플라스틱,.3)}

여기서 모양이 원형인 것을 찾는다면, B는 완전히 만족되는 것이며, A는 중간 정도 만족되는 것이다. 라프집합은 B를 양영역, A를 경계영역에 포함된 객체라 한다.

정의 3.1: $[d_x]$를 d_x를 가진 동치 클래스라 할때, 튜플 $t_i=\{(d_{i1},\mu_{i1}),(d_{i2},\mu_{i2}),...,(d_{im},\mu_{im})\}$와 튜플 $t_k=\{(d_{k1},\mu_{k1}),(d_{k2},\mu_{k2}),...,(d_{km},\mu_{km})\}$는 $[d_{ij}]=[d_{kj}]$이면 중복이다.

퍼지라프 관계에서 속성값이 중복되나 μ값만이 다를 수 있다. 이 경우 퍼지라프 관계에서 합병중 중복 튜플을 제거할 때 가장 높은 μ값의 튜플이 남도록 한다. 이것은 두 퍼지라프집합을 합집합할 때, 최고 소속도가 할당되도록 하는 것이다. 이는 하나의 정보원이 정확한 부분 정보를 갖고 있고, 동일 정보에 대해 다른 정보원이 좀 부정확한 정보를 가질 때, 일반적으로 그 정보가 정확하다고 보기 때문이다.

정의 3.2: X와 Y의 퍼지라프 union X∪Y는 다음과 같은 퍼지 라프 관계 T이다.

$$T = \{t \mid t\in X \text{ or } t\in Y\}$$
$$\mu_T(t)=MAX[\mu_X(t),\mu_Y(t)]$$

즉, 퍼지라프 관계 T는 X와 Y에 있는 모든 튜플을 합병하고 중복 튜플을 제거한 것이다. X가 Y와 같은 속성값을 가질때는 높은(MAX 연산) μ값의 것을 취한다.

<관계 X>

객체	색 깔
A	{(빨강,.7),(노랑,.5),(파랑,.2)(초록,.1)}
B	{(빨강,.5),(노랑,.3),(파랑,.2)}
C	{(빨강,.7),(노랑,.3),(초록,.3)}

<관계 Y>

객체	색 깔
A	{(빨강,.5),(노랑,.2),(파랑,.2)(초록,1.)}
B	{(빨강,.3),(노랑,.1),(파랑,.3)}

<관계 T = X ∪ Y>

객체	색 깔
A	{(빨강,.7),(노랑,.5),(파랑,.2)(초록,.1)}
B	{(빨강,.5),(노랑,.3),(파랑,.3)}
C	{(빨강,.7),(노랑,.3),(초록,.3)}

여기서, $\mu_{T빨강}(B)$ = max(.5,.3) = .5, $\mu_{T노랑}(B)$ = max(.3,.1) = .3, $\mu_{T파랑}(B)$ = max(.2,.3) = .3이다.

정의 3.3: X와 Y의 퍼지라프 intersection X∩Y는 다음과 같은 퍼지 라프 관계 T이다.

$$T = \{t \mid t\in X \text{ and } t\in Y\}$$
$$\mu_T(t) = MIN[\mu_X(t),\mu_Y(t)]$$

즉, 퍼지 라프 관계 T는 X와 Y에 있는 같은 튜플을 합병한 것이다. X가 Y와 다른 소속도를 가질때는 작은(MIN 연산) μ값의 것을 취한다.

<관계 X>

객체	색 깔
A	{(빨강,.7),(노랑,.5),(파랑,.2),(초록,.1)}
B	{(빨강,.3),(노랑,.1),(파랑,.1),(초록,.5)}
C	{(빨강,.7),(노랑,.3),(초록,.3)}

<관계 Y>

객체	색 깔
A	{(빨강,.5),(노랑,.2),(파랑,.2)(초록,.2)}
B	{(빨강,.5),(노랑,.3),(파랑,.2)}

<관계 T = X ∩ Y>

객체	색 깔
A	{(빨강,.5),(노랑,.2),(파랑,.2)(초록,.1)}
B	{(빨강,.3),(노랑,.1),(파랑,.1)}

여기서, $\mu_{T빨강}(A)$ = min(.7,.5) = .5, $\mu_{T노랑}(A)$ = min(.5,.2) = .2, $\mu_{T파랑}(A)$ = min(.2,.2) = .2, $\mu_{T초록}(A)$ = min(.1,.2) = .1이다.

정의 3.4: X와 Y의 퍼지라프 difference X - Y는 다음과 같은 퍼지라프 관계 T이다.

$$T = \{t[(d_{i1},\mu_{i1}),(d_{i2},\mu_{i2}),\dots,(d_{im},\mu_{im})]\in X \mid t[(d_{i1},\mu_{i1}),(d_{i2},\mu_{i2}),\dots,(d_{im},\mu_{im})]\notin Y\} \cup \{t[(d_{i1},\mu_{i1}),(d_{i2},\mu_{i2}),\dots,(d_{im},\mu_{im})]\in X \mid t[(d_{i1},\mu_{i1}),(d_{i2},\mu_{i2}),\dots,(d_{im},\mu_{im})] \in Y \text{ and } \mu_{Xi}>\mu_{Yj}\}$$

퍼지라프 관계 T는 X에는 있고 Y에는 없는 튜플을 가진다. X와 Y 양쪽에 있는 튜플은 X에 있는 μ값이 Y에 있는 것보다 큰 것을 선택한다.

정의 3.5: X로부터의 튜플들의 퍼지라프 selection $\sigma_{A=a}(X)$에 의한 퍼지라프 관계 Y는 다음과 같다.

$$Y = \{t\in X \mid \cup_i[a_i\mu_i] \subseteq \cup_j[b_j\mu_j]\}$$

여기서 $a_i\in a$, $b_j\in t(A)$이고, 소속도는 selection 조건이 AND일때는 MIN($\mu_X(ai)$,$\mu_X(bi)$), OR일때는 MAX($\mu_X(ai)$, $\mu_X(bi)$)이다.

<관계 Legobase>

객체	색 깔	모 양	재 질
A	{(빨강,.5),(노랑,.2),(파랑,.2),(초록,.1)}	{(원형,.5),(사각,.3),(삼각,.2)}	{(나무,.3),(플라스틱,.7)}
B	{(빨강,.5),(노랑,.3),(파랑,.2)}	{(원형,1)}	{(나무,.7),(플라스틱,.3)}
C	{(노랑,.7),(파랑,.2),(초록,.1)}	{(원형,.5),(삼각,.5)}	{(나무,1)}
D	{(노랑,1)}	{(원형,1)}	{(플라스틱,1)}

(예) select 객체 from Legobase
where 색깔 = 노랑 and 모양 = 원형
and 재질 = 플라스틱

과정 : 노랑, 원형, 플라스틱을 가진 튜플은 A, B, D이다. 각각의 소속도는

A = min(.2,.5,.7) = .2

B = min(.3,1,.3) = .3

D = min(1,1,1) = 1

즉, 튜플 D는 색깔이 완전 노랑, 원형, 플라스틱이므로 양영역으로 선정되며, 튜플 A, B는 일부 노랑, 원형, 플라스틱 이므로 경계영역으로 선정된다.

결과 : D, [B, A]

정의 3.5: B로의 X의 퍼지라프 projection $\pi_B(X)$는 다음과 같은 스키마 Y(B)를 가진 퍼지라프 관계 Y이다.

$$Y(B) = \{t(B) \mid t\in X\}$$

정의 3.6: 두개의 관계 X, Y의 퍼지라프 join $X_{<JOIN\ COND>}Y$는 다음과 같은 관계 T이다.

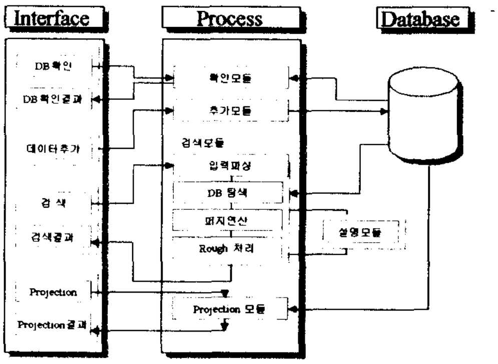
$$T = \{t \mid \exists t_x\in X,\ t_y\in Y \text{ for } t_x\in t(A),\ t_y=t(B)\}, \text{ and where } t_x(A\cap B)=t_y(A\cap B),\ \mu = \min(\mu_x,\mu_y)$$

관계 T는 관계 X의 튜플 x와 관계 Y의 튜플 y에 대해 join 조건을 만족하는 모든 x와 y를 접속해서 만들어지는 튜플로 된 관계이다. <JOIN COND>는 A=B인 여러 조건의 conjunction이다. 조건이 만족된 경우 두 튜플의 소속도중 최소치를 취한다.

4. 실험

3절에서 정의한 연산자중에서 Select와 Projection에 대하여 Visual Basic 5.0, VBCGI, FolkWeb Server, Java Script를 사용하여 퍼지라프 관계모델의 프로토타입을 구현하여 실험해 보고자 한다.

4.1 시스템의 구조



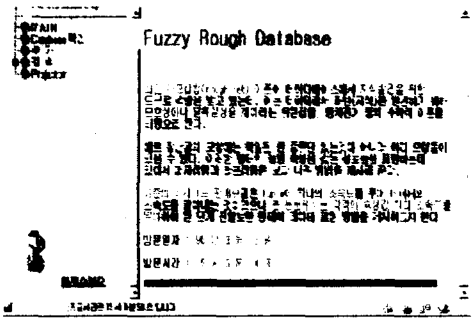
[그림-1]

시스템은 그림-1과 같이 인터페이스 부분, 처리 부분, 그리고 데이터베이스로 구성이 되어 있다. 처리 부분은 다음과 같이 4개의 모듈로 구성된다.

- 확인 모듈: 데이터베이스의 내용을 확인하고자 할 때 데이터베이스의 전부를 출력한다.
- 추가 모듈: 새로운 데이터를 데이터베이스에 추가하는데, 화면을 통하여 직접 입력하면 데이터베이스에 기록된다.
- 검색 모듈: 데이터베이스에서 필요한 튜플을 찾아내는 모듈로서, 인터페이스를 통해 조건 검색문이 들어오면 속성값 및 조건을 파싱한다. 속성값에 따라 데이터베이스를 탐색하면 탐색된 튜플들을 조건에 따라 퍼지 연산을 한다. 연산 결과에 따라 검색 조건에 완전히 일치하는 튜플과 일부 만족하는 튜플을 라프 처리하여 출력한다. 이때 설명 모듈을 사용하여 연산과정을 출력할수도 있다.
- 프로젝션 모듈: 특정 속성만 필요할 때 속성명을 입력하면 그 속성에 대한 프로젝션 결과를 출력한다.

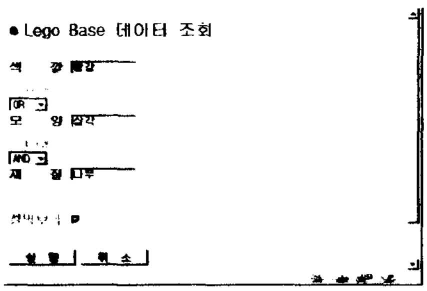
인터페이스 부분은 그림-2와 같이 작업을 메뉴로 선택하도록 한다. <MAIN>은 시작 홈페이지이며, <Database 확인>은 데이터베이스의 내용을 확인할수 있도록 한다. <추가>는 데이터베이스에 새로운 데이터

를 추가할수 있도록 하며, <검색>은 검색 질의를 처리하도록 한다. <Projection>은 특정 속성만 보고자 할 때 사용한다.



[그림-2]

이중 검색 메뉴는 그림-3과 같이 검색할 속성값을 입력하도록 하고 조건은 콤보 박스에서 선택하도록 한다. 출력시 설명이 필요하면 설명보기 체크박스를 체크한다.



[그림-3]

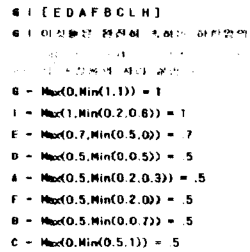
4.2 시스템의 실험

실험에 사용한 관계 데이터베이스는 표-3와 같이 색깔, 모양, 재질 속성을 가진 장난감 데이터베이스이다.

객체	색 깔	모 양	재 질
A	빨강,0.5,노랑,0.2,파랑,0.2,초록,0.1	원형,0.5,사각,0.3,삼각,0.3	나무,0.3,플라스틱,0.7
B	빨강,0.5,노랑,0.3,파랑,0.2	원형,1	나무,0.7,플라스틱,0.3
C	노랑,0.7,파랑,0.2,초록,0.1	원형,0.5,삼각,0.5	나무,1
D	빨강,0.5,파랑,0.2,초록,0.3	사각,1	나무,0.5,플라스틱,0.5
E	노랑,0.3,빨강,0.7	삼각,0.5,사각,0.5	플라스틱,1
F	빨강,0.5,파랑,0.5	사각,0.8,삼각,0.2	플라스틱,1
G	노랑,0.8,파랑,0.2	삼각,1	나무,1
H	파랑,0.2,초록,0.8	원형,0.8,삼각,0.2	나무,0.4,플라스틱,0.6
I	빨강,1	원형,0.3,사각,0.5,삼각,0.3	나무,0.6,플라스틱,0.4
J	노랑,1	원형,1	플라스틱,1
K	파랑,1	원형,0.5,사각,0.5	나무,0.9,플라스틱,0.1
L	초록,1	원형,0.4,사각,0.3,삼각,0.3	나무,1

[표-3]

검색을 위해 그림-3과 같이 입력하면 Web에서 HTML 문서로 입력받아 POST 방식으로 데이터를 넘겨주며 VB CGI 프로그램을 실행시킨다. 실행 결과는 그림-4와 같다. 계산과정이 설명 옵션으로 출력되며, 검색 결과에서 [ ]속에 표시된 것은 경계영역의 객체들인데, 소속도가 높은 것부터 정렬이 되어 있다.



[그림-4]

5. 결론

본 논문에서 제안한 퍼지라프 관계모델은 튜플의 속성값과 관련된 퍼지 소속도를 사용하여 확장한 것이다. 즉, 튜플의 각각의 속성값마다 소속도를 부여하여 데이터 값을 표현했다. 따라서 본 모델은 일반 관계 뿐만 아니라 퍼지 및 라프집합의 모든 특성을 가지고 있으므로 불확실한 데이터를 표현할수 있고, 질의에 대한 양영역의 결과 뿐 아니라 경계영역의 결과도 퍼지 소속도의 순서에 따라 제공할수 있다.

본 연구에서는 우선 속성값의 표현에 이산치의 소속도만 표현했으나 퍼지함수를 사용할수 있도록 확장해야 하며, null값이나 정의가 안된 값도 표현 및 처리할수 있는 연구가 계속되어야 할 것이다.

참고문헌

[1] T. Beaubouef and F.E. Petry, "Fuzzy Set Quantification of Roughness in a Rough Relational Database Model", 1994

[2] T. Beaubouef and F.E. Petry, "A Rough Set Model for Relation Datatbase", 1993

[3] E. H. Mamdani, "Applications of fuzzy algorithms for control of simple dynamic plant," IEEE Proc. 121, pp.1585-1588, 1974

[4] Z. Pawlak, "ROUGH SETS:Theoretical Aspects of Reasoning about Data", Kluwer, 1991

[5] Z. Pawlak, "Rough Sets", International Journal of Computer and Information Science, 11, pp.341-356, 1982

[6] M. Wygralak, "Rough Sets and Fuzzy Sets - Some Remarks on Interrelations," Fuzzy Sets and Systems, 29, pp.241-243, 1989

[7] L. A. Zadeh, "Fuzzy Sets," Inform. anf Control, 8, pp.338-353, 1965

[8] 김도현 외 6명, "핵심 퍼지시스템", 1994

[9] 퍼지기술연구회, "퍼지시스템입문", 1992