퍼지-라프 집합을 이용한 관계 데이터베이스 구성

강 전 근† · 정 환 묵††

요 약

본 논문에서는 의미적으로 접근되어 애매성이 있으며 구별하기 힘든 자료를 효과적으로 관리하기 위한 방안의 하나로 퍼지 집합과 라프 집합을 통합한 퍼지-라프 집합(Fuzzy-Rough Sets : FRS)을 이용, 관계 데이터베이스를 구성하고 구현하였다.

먼저 불완전한 정보를 데이터베이스로 구성, 표현하는 방법에 대하여 일반적 관계 데이터베이스를 확장시킨 퍼지 데이터베이스와 라프 데이터베이스를 간략히 살펴본다. 그리고 퍼지 집합과 라프 집합을 통합한 퍼지-라프 집합을 근거로 퍼지-라프(Fuzzy-Rough : FR)관계 데이터베이스를 구성한 후, 펜티엄 컴퓨터(166Mhz) 상에서 데이터베이스 관리 시스템인 엑세스(access)와 비쥬얼베이직(visual basic)을 도구(tool)로 구현하고 분석하였다.

본 논문에서는 퍼지 집합의 특성과 라프 집합의 특성을 가진 퍼지-라프 집합을 기반으로한 데이터베이스를 구성, 구현함으로서 데이터의 불완전성 감소를 유도하였다.

Design and Implementation of Relational Database Model Using Fuzzy-Rough Sets

Jeon-Geun Kang† · Hwan-Mook Chung††

ABSTRACT

In this paper, for useful administration of the data which have ambiguities meaningfully and are hard to treat, a new relational database model using an integrated fuzzy and rough sets is proposed.

We explain the extended standard relational database model representing fuzzy sets and rough sets relational database one. After proposing Fuzzy-Rough relational database model on the base of integrated Fuzzy and Rough sets, Application of the examples of arithmetic is analyzed through the Access DBMS and the visual basic by composing and representing database based on fuzzy and rough sets which are characterized as fuzzy sets and rough sets on Pentium computer(166Mhz).

This paper was induced to reduce the data incompleteness.

1. 서 론

현실 세계에는 명확히 정의될 수도 없고 애매하며 구별하기가 곤란한 데이터가 다수 존재한다. 이와 같은 데이터를 취급하기 위해서는 지금까지 데이터베이스를 다루어 오던 종래의 방법으로는 곤란하다. 따라서 이들을 다루기 위한 방안으로는 일반적 데이터베이스 개념을 확장시켜 퍼지 이론을 적용한 퍼지 데이터베이스가 있고, 퍼지개념과 유사하게 근사 개념을 이용한 라프 관계 데이터베이스 모델이 제안 되었

† 종신회원 : 영진전문대학 전자계산과 교수
†† 정 회 원 : 대구효성가톨릭대학교 전자정보공학부 교수

논문접수 : 1996년 8월 8일, 심사완료 : 1996년 12월 30일

다[7]. 퍼지 데이터베이스와 라프 데이터베이스의 제안된 모델에 대하여 간략히 살펴보면, 현실 세계에 존재하는 애매한 데이터를 축적, 검색, 처리하기 위해서는 종래의 일반적 데이터베이스 모델로는 불충분하며 모든 속성 값을 고려하여 적당히 계산된 귀속도(degree of membership)를 부여할 수 있는 퍼지개념의 도입 필요성이 제기된다. 이에대한 연구로 Buckles과 Petry등[1][11]은 동치 관계를 퍼지화한 유사 관계(similarly relation)를 도입하여 관계 데이터베이스 모델을 정립하였으며 Umano는 데이터 자체가 갖는 애매성을 가능성 분포로 사용하고 속성 값을 적용할 수 없는 경우에 특별한 값을 정의구역에 추가하여 적용하는 가능성 분포 확장 모델을 발표 하였다[10].

라프 집합에 대한 개념은 1982년 Pawlak[3]에 의하여 정리되었으며 집합론의 근사(approximation)개념을 이용하여 부정확한 데이터 속성에 대한 관계를 형식적으로 정의함으로써 명확히 구별하기가 곤란한, 모호한 데이터의 분류 및 추론을 가능케 한 것이다.

라프 관계 데이터베이스 모델은 일반적 데이터베이스를 기본으로 하여 상한(upper)과 하한(lower)에 대한 근사를 집합(set)의 정리로써 확장한 것인데 대표적인 것으로는 Petry와 Theresa의 라프 집합 관계 데이터베이스 모델이 있다[7].

어떤 사람이 대화중 상대방의 나이를 질의 한 경우 쉰 살쯤 된다고 했을 때, 실제 그의 나이는 50세일 수도 있고 49세나 51세가 될 수 도있는 나이가 된다. 질문자가 수치적으로 정확한 답변을 요구한 것이 아니라면 이들 나이의 수치는 불확실하고 애매성이 있으며 명확히 구별하기 힘든, 모호하면서도 가능성 있는 범주 내의 답변이지만 답으로서 만족해하고 다음 대화로 이어 간다. 또 국적을 문의했을 때, 우리나라 국적을 대한민국이라고 하거나 한국이라는 말로 표현하게 되는데 이 답변은 사실상 동일한 의미임을 알 수 있다. 또 의미적으로 거의 유사한 도랑과 개울 같은 낱말도 있다. 이와 같이 의미적으로 아주 근접되어 유사성이 있거나 구별할 수 없는(Indiscernible) 데이터들이 현실 세계에 다수 존재함을 알 수 있다. 따라서 퍼지 집합과 라프 집합의 특성인 애매하면서도 구별하기 모호한(Indiscernible) 요소를 하나의 공간내에서 관리 해줄 필요성이 제기 된다. 본 논문에서는 이들을 다루기 위한 방안으로서 퍼지 집합과 라프 집합을 통합한 퍼지-라프 집합을 이용하여 애매하고 유사성 있으며 구별하기 모호한 데이터들을 주제로 데이터베이스를 구성하고 연산 예와 함께 구현한다.

1장은 서론이고 2장에서는 퍼지-라프 집합과 불완전 정보의 데이터베이스화에 대한 표현 방법들을, 3장에서는 퍼지-라프 집합의 연산자와 함께 관계 모형을 형식적으로 구성하며, 4장에서는 연산 예와 함께 구현 내용을 제시한 후, 5장에서 결론 및 연구 방향을 기술한다.

2. 퍼지-라프 집합과 불완전 정보의 데이터베이스

퍼지 집합과 라프 집합의 통합된 구조와 불완전 정보의 데이터베이스 표현에 대하여 기술한다.

2.1 퍼지 집합

Zadeh에 의하여 발표된 퍼지 집합을 다음과 같이 정리한다. X는 집합이고 L은 속(Lattice)이라 할 때 L은 폐구간 [0, 1]이 될 수 있다. X에서 퍼지 집합 A는 각 요소 $x \in X$이고 소속 정도(degree of membership) $\mu_A(X) \in L$과 관련된 소속 함수 $\mu_A : X \rightarrow L$에 의하여 응용된다.

A와 B를 X에서 퍼지 집합이라 하면

$A = B$ iff $\mu_A(x) = \mu_B(x)$ for all $x \in X$,
$A \subset B$ iff $\mu_A(x) \leqq \mu_B(x)$ for all $x \in X$,
$C = A \cup B$ iff $\mu_C(x) = \max[\mu_A(x), \mu_B(x)]$ for all $x \in X$.
$D = A \cap B$ iff $\mu_D(x) = \min[\mu_A(x), \mu_B(x)]$ for all $x \in X$.
이 되고

A의 여집합을 A′로 두면 다음과 같다.

$\mu_A{}'(x) = 1 - \mu_A(x)$ for all $x \in X$

2.2 라프 집합

Pawlak에 의하여 소개된 라프 집합을 다음과 같이 정리한다.

$U \neq 0$ 집합일 때 B는 U의 부분 집합으로 불 대수(Boolean algebra) P(U)의 진 부분 대수로 놓으면 이때 (U, B) 쌍을 라프 전집합 (rough uiverse)이라 한다.

$V=(U, B)$를 고정된 라프 전집합(rough universe)으로 두고 R을 다음과 같은 관계로 하면 $A=(A_L, A_U)\in R$ iff $A_L, A_U\in B$, $A_L\in A_U$ 여기서 R의 원소를 라프 집합이라 하며, B의 원소를 정확한 집합(exact sets)이라 한다. 원소 $X\in B$와 같은 원소 $(X, X)\in R$이 동일시 될 때에 정확한 집합은 라프 집합이 된다. 그러나 라프 집합은 정확한 집합이 되지 않는다. 예로 $U\neq 0$이면 (ϕ, u)는 정확하지 못한 집합이 된다[5].

$A=(A_L, A_U)$이고 $B=(B_L, B_U)$를 라프 집합이라고 하면 다음과 같이 정의된다.

$A\cup B=(A_L\cap B_L, A_U\cup B_U)$

$A\cap B=(A_L\cap B_L, Au\cap Bu)$

$A\subset B$ iff $A\cap B=A$

따라서

$A\subset B$ iff $A_L\subset B_L$, $Au\subset Bu$

2.3 퍼지-라프 집합

앞에서 정의된 퍼지 집합과 라프 집합의 특성을 결합한 퍼지-라프 집합은 다음과 같다[6].

U는 집합이고 B는 U의 모든 부분 집합의 Boolean algebra의 Boolean subalgebra이며, L은 속(lattice)이고 X를 라프 집합으로 하면 $X=(X_L, Xu)\in B^2$, $X_L\subset Xu$이 된다.

X에서 퍼지-라프 집합 $A=(A_L, Au)$는 다음 특성을 갖는 소속 함수의 쌍인 $\mu_{AL}:X_L\rightarrow L$ 및 $\mu_{Au}:Xu\rightarrow L$에 의하여 특성화된다.

$\mu_{AL}(x)\leq\mu_{AU}(x)$ for all $x\in Xu$

X에 있는 두 퍼지-라프 집합 $A=(A_L, Au)$와 (B_L, Bu)는 다음과 같이 정의한다.

(1) $A=B$ iff $\mu_{AL}(x)=\mu_{BL}(x)$ for each $x\in X_L$
and $\mu_{AU}(x)=\mu_{BU}(x)$ for each $x\in X_U$

(2) $A\subset B$ iff $\mu_{AL}(x)\leq\mu_{BL}(x)$ for each $x\in X_L$
and $\mu_{AU}(x)\leq\mu_{BU}(x)$ for each $x\in X_U$

(3) $C=A\cup B$ iff $\mu_{CL}(x)=\max[\mu_{AL}(x), \mu_{BL}(x)]$ for all $x\in X_L$, $\mu_{CU}(x)=\max[\mu_{AU}(x), \mu_{BU}(x)]$ for all $x\in X_L$

(4) $D=A\cap B$ iff $\mu_{DL}(x)=\min[\mu_{AL}(x), \mu_{BL}(x)]$ for all $x\in X_L$
$\mu_{DU}(x)=\min[\mu_{AU}(x), \mu_{BU}(x)]$ for all $x\in X_U$

2.4 불완전 정보의 데이터베이스 표현

불완전 정보에 대한 데이터베이스 표현은 널(null) 값에 기초를 둔 접근 방법[4]으로는 효과적이지 못하다. 따라서 불완전 정보를 다루기 위한 여러 방법들이 검토되었는데 그중 애매한 데이터를 적절히 표현하는 방법으로 퍼지 개념을 도입한 퍼지 데이터베이스가 있다.

Petry의 퍼지 데이터베이스를 살펴보면 일반적 관계 모델을 약간 확장하고 여기에 동치관계를 퍼지화한 유사 관계(similarity relation)를 도입, 관계 대수로 불리는 관계 모델에 따른 연산을 수행하도록 하였다. (그림 1)은 이 3요소를 나타낸다.

(a) 퍼지 집합

$A=\{a_1, a_2, a_3, a_4, a_5\}$

$B=\{b_1, b_2, b_3, b_4\}$

$C=\{c_1, c_2, c_3, c_4\}$

(b) 관계(Relation)

R_1

A	B
a_1	b_2
a_3	b_1
a_5	b_4
a_1	b_3

(c) 유사 관계

	a_1	a_2	a_3	a_4	a_5
a_1	1	0.6	0.3	0.6	0.5
a_2	0.6	1	0.3	0.7	0.5
a_3	0.3	0.3	1	0.3	0.3
a_4	0.6	0.7	0.3	1	0.5
a_5	0.5	0.5	0.5	0.5	1

(그림 1) Petry의 퍼지 데이터베이스 예
(Fig. 1) Example of Fuzzy Database made by petry

이와 같은 퍼지 데이터베이스에 대하여 $R_2 \leftarrow$(project(R_1 over A) with Level(A)=0.4)를 실행하는 경우 R_2와 같은 퍼지 관계를 얻을 수 있다.

R_2

A
$\{a_1,a_5\}$
a_3

이 결과는 관계 R_2에 2개 이상의 요소를 포함하게 된다. 이는 일반적인 관계 모델로는 나타낼 수 없다.

Pawlak에 의하여 제시된 라프 집합은 구별하기 모호한 근사 개념을 표현하는데 효과적이다. 예를 들어 다음과 같은 관계 R_3에서 Ind(Indiscernibility)번호의

R_3

도메인	속성값	Ind번호
country	korea	101
country	hankook	101
country	us	201
country	usa	201
country	mexico	301
country	japan	401
feature	forest	501
feature	woods	501
feature	jungle	501

도메인(Domain)으로 표시한 101은 {korea, hankook}이 속성인 country의 도메인이고 Ind번호 501은 속성 feature의 도메인으로 {forest, woods, jungle}이 동일한 의미를 갖는 것으로 나타낸 예이다. 다음으로 관계 R_4는 모든 튜플이 라프 집합의 하한근사로 이루어 구성되어 졌고 { }의 요소는 복수의 속성 값으로 구성되어 동치 류를 나타낸다.

R_4

ID번호	국 가	특 징
ID1	korea	{river,beach}
ID2	usa	forest
ID3	mexico	{sand,road}
ID4	japan	sand
ID5	{korea,us}	{river,lake}
ID6	hankook	river

관계 R_4로부터 다음과 같은 조회를 하고 결과를 살펴보자.

```
SELECT X.ID
FROM R4
WHERE COUNTRY="korea" and feature="river"
```

결과를 분석해 보면, 명확한 정보인 ID6와 라프 집합의 경계에 속하는 근사 정보인 {ID1, ID5}를 얻을 수 있다. 여기서 country와 feature속성의 도메인 값은 동치 클래스로 다음과 같이 나타낼 수 있다.

{[korea, hankook], [usa, us], [mexico], [japan]},
{[river, lake], [sand, beach], [road], [forest]}

불완전한 정보에는 살펴본 바와 같이 퍼지 개념을 이용, 소속 함수에 의존하는 확률적인 요소와 상한과 하한의 근사 개념을 도입, 경계 영역 관리가 필요한 라프적 요소들이 있음을 알 수 있으며 이들은 때로는 다음 예와 같이 하나의 공간에서 공존되어 관리할 필요가 있다.

R_5

이름	나이	국적	Ind번호
kim	{25,26}	korea	101
park	17	hankook	101
tom	{30,32}	us	201
charles	47	usa	201
david	60	america	201

R_5테이블에서 park의 17세는 명확한 나이이나 kim의 나이는 25세인지 26세인지 불확실하게 추정되는 나이임을 의미한다. Ind번호로 표시된 것은 속성인 국적의 구별할 수 없는(Indiscernibility) 요소로 {[korea, hankook], [us, usa, america]}와 같이 동치 클래스의 집합을 나타낸다.

3. FRS관계 데이터베이스 구성

2장에서 서술된 퍼지-라프(Fuzzy-Rough)집합에 근거하여 퍼지적 요소와 라프적 요소로 이루어진 FRS 관계 데이터베이스를 다음과 같이 구성한다.

3.1 FRS관계 데이터베이스

일반적 관계 데이터베이스를 기초로 유사 관계를

도입, 동치 관계를 퍼지화한 특성과 라프적 요소를 의미하는 구별할 수 없는 요소(Indiscernibility)를 취합, FRS관계 데이터베이스를 다음과 같이 구성한다 [6], [7].

1) 항상 원자 값만 취하는 속성은 원자값 속성이다.

2) 제1차 키(Primary Key)는 원자값 속성으로 구성된다.

3) 속성(Attribute)이 퍼지 성질이면 속성 값도 퍼지 성질을 갖는 도메인(Domain)으로 구성된다.

4) 속성이 라프 성질을 가지면 속성 값도 라프 성질을 갖는 도메인으로 구성된다.

5) 속성이 퍼지 성질을 가질때 속성 값은 유사성을 가진 동치관계 요소의 집합으로 구성된다.

6) 속성이 라프 성질을 갖으면 속성 값은 근사적이며 구별할 수 없는(Indiscernibility) 요소의 집합으로 구성된다.

7) 속성이 퍼지-라프 집합적 성질을 가지면 속성 값도 퍼지-라프 집합적 성질을 갖는 도메인으로 구성된다.

8) FRS관계 데이터베이스는 다음의 특성을 갖는다

(1) 중복된 튜플(tuples)은 존재하지 않는다.

(2) 튜플의 순서는 바뀌어도 상관없다.

(3) 한 관계를 구성하는 속성 사이에는 순서가 없다.

(4) 모든 속성 값은 원자값 이상이다.

9) FRS관계는 다음의 무결성 제약 조건을 갖는다.

(1) 객체(entity)의 무결성: FRS관계에는 적어도 하나의 객체 속성으로 구성되나 널 혹은 널 값은 가질 수 없다.

(2) 참조 무결성: FRS관계의 속성 A는 제1차 키 도메인(domain)에서 정의한다. 이때 대응이 되는 외래키(foreign key)값, FKn를 관계 R가 가지고 있다하면 FKn의 값은 기본 키의 어떤 값과 같든지 아니면 널이어야 한다.

3.2 FRS관계 데이터베이스의 기본 연산

일반적 관계 데이타베이스 연산자를 FRS관계 데이터베이스와 퍼지-라프 집합을 참조하여 FRS관계 데이터베이스의 연산자로 확장시켜 적용한다. 참고로 불완전 정보에 대한 관계(Relation)의 상관성을 살펴보면 관계의 술어(Predicate)의 참값(truth value)으로 두 가지 형태의 진리를 생각하게 되는데 하나는 술어 P를 정적으로 만족시키는 값 P_T가 있고 둘째는 P를 만족시킬 가능성이 있는 값(maybe true)으로 P_M을 고려할 수 있으며 이들의 관계는 $P_T \subseteq P_M$으로 나타낼 수 있다. 그리하여 현실 세계에서 연산된 결과는 P_T와 P_M사이에 있다고 볼 수 있으며 이는 현실 세계의 하한(P_T)과 상한(P_M)으로 해석된다[8]. 그러나 FR값은 퍼지-라프 집합 정의에 의하여 상한과 하한이 퍼지-라프 집합 정의 범위 내에서 동일한 처리 범위를 갖는다.

다음의 X와 Y는 퍼지-라프(FR) 관계(Relations)을 나타내며 관계 $X(A_1, A_2, A_3, \ldots, A_n)$와 $Y(B_1, B_2, B_3, \ldots, B_n)$는 합병 가능(Union-compatible)하여야 한다. 여기서 R_L은 FR의 성질을 가진 낮은 접근(근사 및 유사)이고 Ru는 FR의 성질을 가진 높은 접근(근사 및 유사)로 표시된다.

UNION.

X와 Y를 합병 가능한 FR관계로 하면

[정의 1] X와 Y의 유니온, X∪Y는 X또는 Y에 속하는 FR튜플(tuples) t로 구성되는 관계로 다음과 같이 T로 나타낼 수 있다.

$$R_LT = \{t \in R_LX \cup R_LY\} \text{ and } R_uT = \{t \in R_uX \cup R_uY\}$$

낮은 접근의 결과 FR관계 T는 낮은 접근의 X와 Y의 구성원들인 튜플을 포함하고 높은 접근의 결과 FR관계 T는 높은 접근의 X와 Y의 구성원들인 튜플을 포함한다.

INTERSECTION.

FR인터섹션(Intersection)은 FR적 성질을 가진 근사 및 유사적인 것으로 정의된다.

[정의 2] X와 Y의 인터섹션, X∩Y는 FR관계 T로 나타낼 수 있다.

$$R_LT = \{t \in R_LX \cap R_LY\} \text{ and } R_uT = \{t \in R_uX \cap R_uY\}$$

낮은 접근의 결과 FR관계 T_L는 낮은 접근의 X와 Y의 구성원들을 중복되는 영역 안에 있으면 모두 포함된다. 높은 결과의 FR관계 T_U도 같은 개념을 갖는다.

DIFFERENCE.

X와 Y를 합병 가능한 FR관계로 하면

[정의 3] FR디퍼런스(Difference), X-Y에서 X와 Y 사이는 FR튜플 t로 구성되는 관계로 다음과 같이 T로 나타낼 수 있다.

$R_LT=\{t\in R_LX \mid t\notin R_LY\}$ and $RuT=\{t\in RuX \mid t\notin RuY\}$

낮은 접근의 결과 T와 높은 접근의 결과 T는 중복 성이 없는 튜플 Y를 제거한 것이다.

PROJECTION.

X를 FR관계로 스키머 A로 두고 B는 A의 부분집합이라 하면, B≤A가 있을 때 관계 R의 속성 B에 대한 프로젝션(Projection) R[B]는 다음과 같이 정의한다.

[정의 4] X상의 B의 FR프로젝션을 $\pi_B(X)$라 할 때 다음과 같이 나타낼 수 있다. $R(B)=\{t(B) \mid t\in X\}$결과 관계에서 중복된 튜플은 하나만 남기고 모두 제거한다.

SELECTION.

퍼지-라프적 데이타베이스를 위한 셀렉션(Selection) 연산자 δ는 독립 변수로서 FR관계 X를 갖는 단항 연산자이며 특정한 속성을 위한 값의 기초 위에서 선택된 X의 튜플들의 부분 집합을 함유한다.

예로 $\delta_{A=a}(x)$에서 a의 값과 동일한 속성 A는 x로부터 튜플들을 선택한다. 정확하게 표현하면 그 값은 [a]가 된다.

R을 관계 데이터베이스의 스키머(Schema)로 놓으면, X는 스키머상의 FR관계이고 A는 R의 속성일 때 a_i, $b_j\in dom(A)$로서 $a=\{a_i\}$가 된다. U_X는 X상의 전체 유니온(Union)을 뜻한다.

[정의 5] FR셀렉션(Selection)이 $\delta_{A=a}(X)$일 때 X로부터의 튜플은 FR관계 Y가 스키머 X에서 $a_i\in a$, $b_j\in t(A)$의 조건을 가지는 것으로 다음과 같다.

$R_LY=\{t\in X \mid U_i[a_i]\subseteq U_j[b_j]\}$ and $RuY=\{t\in X \mid U_i[a_i]B\subseteq U_j[b_j]\}$

낮은 접근과 높은 접근에서 모두 a의 멤버들은 속성 A값의 부분 집합이 되는 튜플들을 함유한다.

JOIN.

조인(join)연산자는 2항 연산자로 두개의 관계로부터 관계된 튜플을 얻게 되는데 튜플을 결합하기 위해서는 일반적으로 θ-조인을 고려하게 되며 θ는 $\{=, \neq, <, >=, >, >=\}$중의 하나이다.

[정의 6] FR조인, X ∞〈조인 조건〉 Y는 관계 X와 Y가 관계 T(C1, C2, ..., Cm +n)가 되는데 이것은 다음과 같이 정의된다.

$T=\{t \mid \exists t_x\in X, t_y\in Y$ for $t_x=t(A), t_y=t(B)\}$, and where
$t_x(A\cap B)\subseteq t_y(A\cap B)$ or $t_Y(A\cap B)\subseteq t_x(A\cap B)$, for R_LT
$t_x(A\cap B)\subseteq t_Y(A\cap B)$ or $t_Y(A\cap B)\subseteq t_X(A\cap B)$, for R_UT

$X(A_1, A_2, ..., A_m)$과 $y(B_1, B_2, ..., B_n)$에서 m과 n은 속성이고 FR관계이며 AB=C인 형태로 조인된 결과를 표시하는데 결과 또한 T의 스키머가 된다.

3.3 FRS관계 데이터베이스 연산자의 성질

관계 대수(relational algebra)는 관계(relations)을 처리하기 위한 연산자(operators)의 집합으로서 많은 특성이 있으며 각 연산자의 피연산자는 모두 관계이고 연산 결과 또한 관계(Relation)라는 특성을 가지고 있으며 이것은 폐쇄 성질(closure property)과 같다. FR관계 데이터베이스에서 사전에 정의된 FR관계 연산자에 의하여 적용된 결과 또한 FR관계가 된다. 셀렉트(select) 연산자를 이용하여 {∪, ∩, −}의 상태 특성을 나타낸다.

$r\in\{\cup, \cap, -\}$로 두고 X와 Y는 두개의 관계 일 때 $\delta_{A=a}(X\gamma Y)=\delta_{A=a}(X)\gamma\delta_{A=a}(Y)$는 연산자 γ에 대한 상태 특성이다.

증명.

Let $\gamma=\cup$.
$=\delta_{A=a}(T)$
$=\delta_{A=a}(X\cup Y)$ where
$R_LT=\{t \mid t\in R_LX$ or $t\in R_LY\}$ and
$R_UT=\{t \mid t\in R_UX$ or $t\in R_UY\}$
$=T'$ where
$R_LT'=\{t'\in\{t \mid t\in R_LX$ or $t\in R_LY\} \mid U_i[a_i]\subseteq U_j[b_j]\}$, $a_i\in a$, $b_j\in t'(A)$, and
$R_UT'=\{t'\in\{t \mid t\in R_UX$ or $t\in R_UY\} \mid U_i[a_i]\subseteq U_j[b_j]\}$, $a_i\in a$, $b_j\in t'(A)$
$=Q$ where

$R_LQ=\{t|t\in X \text{ or } U_i[a_i]\subseteq U_j[b_j]\}\cup\{t|t\in Y \text{ or } U_i[a_i]\subseteq U_j[b_j]\}$

$R_UQ=\{t|t\in X \text{ or } U_i[a_i]\subseteq U_j[b_j]\}\cup\{t|t\in Y \text{ or } U_i[a_i]\subseteq U_j[b_j]\}$

$=\delta_{A=a}(\{t|t\in X\})\cup\delta_{A=a}(\{t|t\in Y\})$

$=\delta_{A=a}(X)\cup\delta_{A=a}(Y)$

Let $\gamma=\cap$.

$=\delta_{A=a}(T)$

$=\delta_{A=a}(X\cap Y)$ where

$R_LT=\{t|t\in R_LX \text{ and } t\in R_LY\}$ and

$R_UT=\{t|t\in R_UX \text{ and } t\in R_UY\}$

$=T'$ where

$R_LT'=\{t'\in\{t|t\in R_LX \text{ and } t\in R_LY\}|U_i[a_i]\subseteq U_j[b_j]\}$, $a_i\in a$, $b_j\in t'(A)$, and

$R_UT'=\{t'\in\{t|t\in R_UX \text{ and } t\in R_UY\}|U_i[a_i]\subseteq U_j[b_j]\}$, $a_i\in a$, $b_j\in t'(A)$

$=Q$ where

$R_LQ=\{t|t\in X \text{ and } U_i[a_i]\subseteq U_j[b_j]\}\cap\{t|t\in Y \text{ and } U_i[a_i]\subseteq U_j[b_j]\}$

$R_UQ=\{t|t\in X \text{ and } U_i[a_i]\subseteq U_j[b_j]\}\cap\{t|t\in Y \text{ and } U_i[a_i]\subseteq U_j[b_j]\}$

$=\delta_{A=a}(\{t|t\in X\})\cap\delta_{A=a}(\{t|t\in Y\})$

$=\delta_{A=a}(X)\cap\delta_{A=a}(Y)$

Let $\gamma=-$

$=\delta_{A=a}(T)$

$=\delta_{A=a}(X-Y)$ where

$R_LT=\{t|t\in R_LX|t\notin R_LY\}$ and

$R_UT=\{t|t\in R_UX|t\notin R_UY\}$

$=T'$ where

$R_LT'=\{t'\in\{t|t\in R_LX \text{ and } t\notin R_LY\}|U_i[a_i]\subseteq U_j[b_j]\}$, $a_i\in a$, $b_j\in t'(A)$

$R_UT'=\{t'\in\{t|t\in R_UX \text{ and } t\notin R_UY\}|U_i[a_i]\subseteq U_j[b_j]\}$, $a_i\in a$, $b_j\in t'(A)$

$=Q$ where

$R_LQ=\{t\in X \text{ and } U_i[a_i]\subseteq U_j[b_j]\}\cap R_LQ=\{t\notin Y \text{ and } U_i[a_i]\subseteq U_j[b_j]\}$

$R_UQ=\{t\in X \text{ and } U_i[a_i]\subseteq U_j[b_j]\}\cap R_UQ=\{t\notin Y \text{ and } U_i[a_i]\subseteq U_j[b_j]\}$

$=\delta_{A=a}(\{t|t\in X, t\notin Y\})$

$=\delta_{A=a}(X)-\delta_{A=a}(Y)$

4. 연산 예와 구현

위에서 정의한 FRS관계 데이터베이스를 참조, FRS 관계 데이터베이스 연산자의 사용을 예로서 기술한다.

4.1 연산 예

〈표 1〉의 표현 테이블에서 R과 S는 관계 테이블이고 R에서 속성 A는 제1차 키(First Key)를 의미하며 B는 애매한 정보로서 제삼자에게 어떤 사람의 나이를 물었을 때 어떤 사람의 나이가 20세인지 21세인지 불확실하게 추정되는 답변의 경우의 예와 같고 C는 구별하기 모호한(Indiscernible) 도메인들로 구성되어 있음을 나타낸다.

IND는 속성C의도메인들의 IND(Indiscernible)클래스 번호를 의미 한다.

관계 S의 속성 D, E, F, IND는 R의 속성 A, B, C, IND와 같은 유형의 도메인 속성으로 구성되어 있다. FRS정의의 특성을 셀렉션(Selection)의 예를 통해 살펴보면 2)의 a에서 술어(Predicate)의 경우 b의 속성값 중 최대 값과 비교되며 b에서 술어는 속성 값 중 최저 값과 비교된다. 또 셀렉션의 예 (c)에서 IND한 개념을 볼 수 있으며 3)의 θ-조인에서는 퍼지값과 IND한 라프 연산 예를 함께 살펴볼 수 있다.

〈표 1〉 표현테이블

〈Table 1〉 Represention table

R(A,B,C)

A	B	C	IND
a_1	{20,21}	korea	101
a_2	31	hankook	101
a_3	52	us	201
a_4	{55,57}	usa	201
a_5	{60,63}	english	301

S(D,E,F)

D	E	F	IND
a_1	{20,21}	hankook	101
a_2	31	korea	101
a_3	54	usa	201
a_4	{57,59}	usa	201
a_5	{60,64}	japan	301

1)Projection

$\pi_{BC}(R)$

B	C
{20,21}	korea
31	hankook
52	us
{55,57}	usa
{60,63}	english

2)Selection

(a) $\delta_{(B\ <\ 56)}(R)$

A	B	C
a_1	{20,21}	korea
a_2	31	hankook
a_3	52	us

(b) $\delta_{(B\ >\ 54)}(R)$

A	B	C
a_4	{55,57}	usa
a_5	{60,61}	english

(c) $\delta_{(C\ =\ 'korea')}(R)$

A	B	C
a1	{20,21}	korea
a2	31	hankook

3) θ 조인

(a) R $_{[C=F]}$ S

A	B	C	D	F
a1	20,21	korea	a1	20,21
a1	20,21	korea	a2	31
a2	31	hankook	a1	20,21
a2	31	hankook	a2	31
a3	52	us	a3	54
a3	52	us	a4	57,59
a4	55,57	usa	a3	54
a4	55,57	usa	a4	57,59

(b) R $_{[B=E]}$ S

A	B	C	D	E	F
a_1	{20,21}	korea	a_1	{20,21}	hankook
a_2	31	hankook	a_2	31	korea

(c) R $_{[B<E]}$ S

A	B	C	D	E	F
a_1	{20,21}	korea	a_2	31	korea
a_1	{20,21}	korea	a_3	54	usa
a_1	{20,21}	korea	a_4	{57,59}	usa
a_1	{20,21}	korea	a_5	{60,64}	japan
a_2	31	hankook	a_3	54	usa
a_2	31	hankook	a_4	{57,59}	usa
a_2	31	hankook	a_5	{60,64}	japan
a_3	52	us	a_3	54	usa
a_3	52	us	a_4	{57,59}	usa
a_3	52	us	a_5	{60,64}	japan
a_4	{55,57}	usa	a_5	{60,64}	japan

4) R U S

a_1	{20,21}	korea
a_2	31	hankook
a_3	52	us
a_4	{55,57}	usa
a_5	{60,63}	english
a_3	54	usa
a_4	{57,59}	usa
a_5	{60,64}	japan

5) R - S

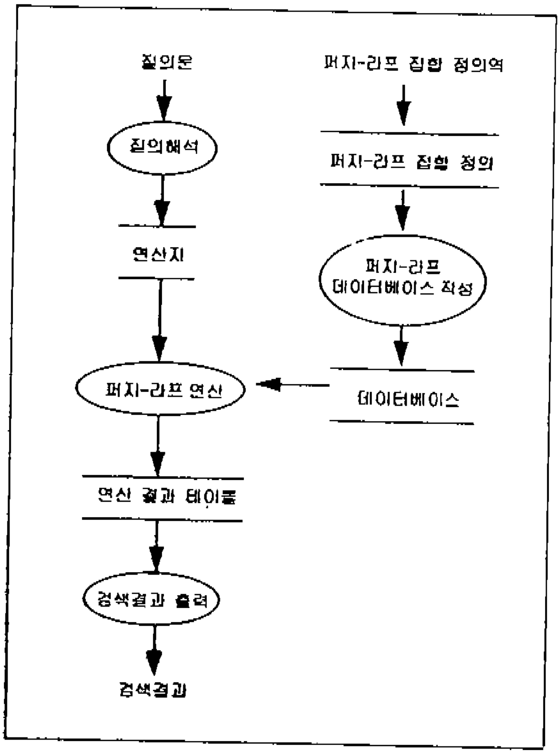
a_3	52	us
a_4	{55,57}	usa
a_5	{60,63}	english

4.2 구 현

연산 예를 토대로 펜티엄(166 Mhz) 상에서 엑세스 데이터베이스 관리 시스템과 비쥬얼베이직 언어를 이용, FRS데이타베이스 질의 시스템을 작성하였다.

(그림 2)는 구현된 FRS데이터베이스 질의 시스템 흐름도 이고 (그림 3)의 1)과 2)는 질의 처리된 전체 시스템 연산 예중 유니온(union)과 조인(join)의 한 예를 나타낸다.

구현된 시스템은 응답 속도 면에서도 질의문을 발행한 후 기다림이 없을 만큼 좋은 반응을 보였다.



(그림 2) 퍼지-라프 질의 시스템 흐름도
(Fig. 2) Fuzzy-Rough query system flow

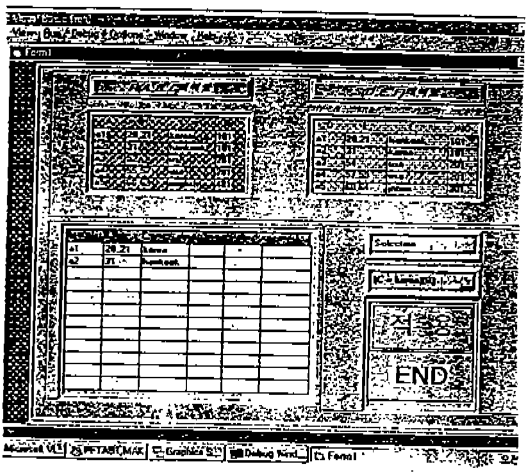
5. 결 론

실세계에는 의미적으로 근접되어 애매하면서도 수치적으로 구별하기 힘든 데이터들이 다수 존재하는데 이들도 하나의 공간 내에서 관리해 줄 필요성이 있다. 그리하여 애매함을 주제로 하고있는 퍼지 집합 개념과 구별하기 힘들고 모호한(Indiscernible) 자료를 근사 개념으로 처리하고 있는 라프 집합 개념, 또 양자의 특성을 대상으로한 통합된 공간인 퍼지-라프 집합(FRS)을 살펴보고 FRS를 기반으로 하여 데이터베이스를 구성, 구현해 봄으로서 불완전한 데이터를 관리하는 하나의 방안을 제시, 불완전성 감소를 유도하였다.

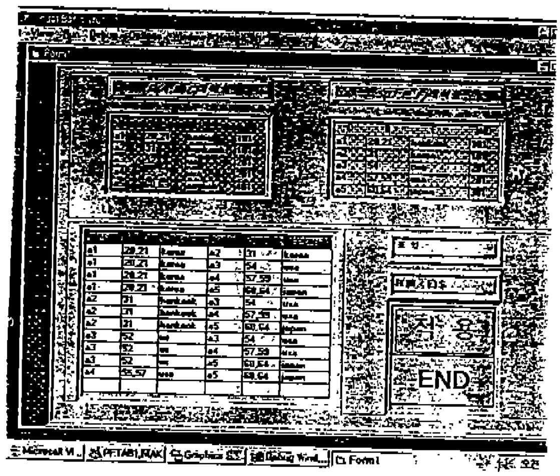
향후 퍼지-라프 집합 개념을 이용한 모호성 감소에 대한 추가적 연구의 필요성과 함께 다양한 응용 분야가 기대된다.

참 고 문 헌

[1] Buckles, B.P and F.E.Petry, "Uncertainty models in information and database systems", Journal of Info. Science, pp. 77-87, 1985.
[2] E.F. Codd, "A Relational model of data for large



1) 셀렉션 [$\delta_{(C='korea')}(R)$]



2) θ-조인 [$R_{[B<E]}S$]

(그림 3) 구현 예
(Fig. 3) Example of Implementation

shared data banks", Comm. ACM, pp. 377-387, 1970.
[3] Z. pawlak, "Rough sets", Int. J. Computer and Info. science, 1982.
[4] J. Grant, "Null Values in a Relational Database", Infor. process, pp. 156-157, 1977.
[5] Zdzisiaw Pawlak, "Rough Sets and Fuzzy Sets", Fuzzy sets and Systems, Elsevier Science pub. V.(North-holland), pp. 99-102, 1985.
[6] S. manda, S.Majumdar, "Fuzzy Rough Sets" Fuzzy Sets and Systems, Elsevier science Pub. V.(North-Holland), pp. 157-160, 1992.
[7] F. E. Petry and Theresa Beaubouef, "A Rough Set Model for Relational Databases", LAQSF-GF-13, British Computer Society, pp. 101-107, 1995.
[8] S. Park "Extended Relational Algebra and Non-truth-Functionality in the Incomplete Relational Model", J. Kiss Vol. 12-3-4, pp. 204-214, Aug. 1985.
[9] J. G. Kang "A Study of conflict prevention for the design of distributed Database", Research review Vol. 9, Yeungjin junior college, pp. 211-216, 1987.
[10] Motohide Umano, "A Fuzzy database system", Infor. process, pp. 667-675, 1982.
[11] Frederick E. Petry and Billy P. Buckles "A Fuzzy representation of data for Relational Databases", Infor. Process, pp. 660-666, 1980.

강 전 근

1977년 2월 동국대학교 전자공학과 졸업
1985년 2월 한양대학교 전자계산학전공(공학석사)
1994년 8월 대구효성가톨릭대학교 전자계산과 박사과정 수료
1979년 6월~1985년 2월 한국전력공사 정보처리처 근무
1985년 3월~현재 대구 영진전문대학 전자계산과 교수
관심분야: 인공지능, 데이타베이스

정 환 묵

1972년 2월 한양대학교 전자공학과 졸업
1987년 2월 인하대학교 박사과정수료(이학박사)
1986년 12월~1987년 12월 일본 동경대학 정보과학과 객원연구원
1995년 2월~1996년 2월 일본 명치대학교 정보공학과 객원교수
1984년 3월~현재 대구효성가톨릭대학교 전자정보공학부 교수
관심분야: 인공지능, 퍼지논리, 다치논리, 지능시스템공학