

# 러프 집합을 이용한 관계데이터베이스 모델의 구성 및 해석

† 정구범\*

‡ 정환목

†영동전문대학 전자계산과  
강원도 명주군 성산면 금산리 11

‡대구효성가톨릭대학교 전자·정보공학부  
경북 경산군 하양읍 금락1리

## Constructions of Relational Database Model Using Rough Sets and Its Analysis

†Gubum Chung\*

‡Hwanmook Chung

†Dept. of Computer Science,  
Youngdong Junior College

‡Faculty of Electronic & Information Engineering,  
Catholic Univ. of Taegu-Hyosung

### Abstract

In this paper, we construct rough relational database model using approximation concepts of rough set. Also, we analyze the relation between objects, attributes and attribute values and, propose the method that can generate flexible retrieval results.

### 1. 서론

데이터베이스는 이미 일반화된 관계 데이터베이스 모델과 더불어 객체지향 데이터베이스, 멀티미디어 데이터베이스 등의 새로운 패러다임으로 진전되고 있고, 데이터베이스에 기억된 데이터의 종류, 데이터 양의 변화가 다양하여 검색에 대한 추상성이 높은 형식의 출력이 강력히 요구되고 있다.

실세계 환경의 다양한 변화와 사용자의 데이터베이스 시스템에 대한 기대치 상승에 따라 데이터베이스 내에 기억되어 있는 데이터로부터 속성 및 속성 값을 적절히 선택하여 그들의 조합 중에서 특징과 규칙성을 찾아서 보다 추상화된 정보의 추출을 가능하게 하는 고도의 사용자 인터페이스를 구축하는 것이 매우 절실히 요구된다. 각 속성, 속성 값 사이에 성립하는 규칙, 제약, 규칙성 등이 전형적인 지식으로 나타난다.

본 논문에서는 러프 집합의 근사 개념을 이용하여 러프 데이터베이스 모델을 구성하고, 데이터베이스의 대상, 속성, 속성 값의 상호 관계를 해석하여 질의에 대한 유연한 검색 결과를 생성할 수 있는 방법을 제안한다.

### 2. 관계데이터베이스

1970년 코드에 의해 제안된 관계데이터베이스(Relational Database : RDB) 모델은 우리에게 익숙해 있던 테이블의 형태로 데이터를 관리하기 때문에 이용하기 쉽고 또한, 데이터 관리 측면에서도 효과적이라는 장점 때문에 상용화되고 있는 모델 중의 하나이다. 그러나, 전형적인 RDB는 명확한 질의를 통해 명확한 검색 결과를 요구한다는 단점을 극복하지 못하기 때문에 최근에는 퍼지 이론, 신경망, 러프 집합(rough set) 등의 인공 지능 기법을 이용하여 기존의 RDB를 확장함으로써 불확실한 상황 하에서의 효과적인 정보 검색을 위한 노력이 경주되고 있다[2,6].

#### 2.1 퍼지 관계데이터베이스

기존의 RDB는 모호성이 없는 분명한 질의를 통해 명확한 결과를 검색한다. 그러나, 현실 세계에 존재하는

각종 데이터와 질의는 많은 불확실성을 내포하고 있기 때문에 기존의 RDB를 그대로 사용한다면 이러한 모호성의 처리는 불가능하게 된다.

퍼지 데이터베이스는 데이터베이스 내의 모든 속성 값들에 전문가에 의해 적당히 계산된 귀속도를 부여하고 질의에 포함된 모호성은 퍼지 언어 변수를 이용하여 처리를 함으로써 질의 조건에 근사적으로 만족하는 결과를 생성하도록 하고 있다. 또한, 버클과 페트리[5,6] 등은 동치 관계를 퍼지화한 유사 관계(similarity relation)를 도입하여 관계 대수로 불리우는 퍼지 관계 데이터베이스 모델을 정립하였으며, 우마노[4]는 데이터 자체가 갖는 모호성을 가능성 분포로 표현한 가능성 분포 모델을 제안하였고 아울러, 속성 값을 가능성 분포로 사용하고 속성 값을 적용할 수 없는 경우에는 특별한 값을 정의구역에 추가하여 확장된 가능성 분포 모델을 제안하였다.

#### 2.2 러프 관계데이터베이스

러프 집합에 대한 개념은 1982년 폴락[3]에 의해 제안되었으며 집합론의 근사(approximation) 개념을 이용하여 부정확한 데이터의 속성에 대한 관계를 형식적으로 정의함으로써 모호한 데이터의 분류 및 추론을 가능하게 한 것인데 명확한 사실에 대하여 지식의 분류 패턴과 각 데이터 속성과의 관계, 불확실성과 모호성을 근사적으로 표현하는데 효과적이라는 것을 입증했다[1].

러프 관계데이터베이스 모델은 기존의 RDB 모델을 기본으로하여 상한 근사(upper approximation)와 하한 근사(lower approximation)를 정의하고 이와 같은 근사 집합을 이용하여 유사성 분류를 통해 데이터베이스의 질의 검색에 이용한 모델이고 전형적인 것으로는 페트리와 테레사 등[2]에 의한 관계데이터베이스 러프 집합 모델이 있다.

### 3. 정보 시스템(관계데이터베이스)

#### 3.1 정보 시스템 모델

일반적인 정보(관계 데이터베이스) 시스템은 다음과

같은 원소를 갖는 순서 쌍으로 나타내며 이를 표로 나타내면 [표 1]과 같다.

$$S = (X, Q, V, P)$$

- 단, X : 대상  $x_1, x_2 \dots$  들의 유한 집합
- Q : 속성  $q_1, q_2 \dots$  들의 유한 집합
- V : 속성 값  $\alpha, \beta, \gamma \dots$  들의 유한 집합
- P :  $X \times Q \rightarrow V_{q_i}$ 의 속성 값 기술 함수

[표 1] 정보 시스템의 테이블 구성

대상	속성 집합			
	$q_1$	$q_2$	...	$q_n$
$x_1$	$\alpha$	$\beta$	...	$\gamma$
$x_2$	$\alpha$	$\alpha$	...	$\beta$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_n$	$\beta$	$\alpha$	...	$\beta$

단,  $\alpha, \beta, \gamma \in V_{q_i} (i = 1, 2, \dots, n)$

### 3.2 정보 시스템 모델의 해석

속성 값을 기술하는 함수  $P_X$ 는 다음과 같이 정의한다.

[정의 1]

$$P_X : Q \rightarrow V ; P_X(q) = P(x, q)$$

또한, 대상  $x_i$ 와  $x_j$ 가 어떤 속성 집합 Q에 관하여 식별할 수 없는 대상들임을 나타내는 식별 불능 관계(indiscernibility relation)  $ind(Q)$ 를 다음과 같이 정의한다.

[정의 2]

$$\begin{aligned} P(x_i, Q) &= P(x_j, Q) \\ (x_i, x_j) &\in ind(Q) \\ R &= ind(Q) \end{aligned}$$

따라서,  $x_i, x_j$ 가  $P \subset Q$ 인 P에 대하여 식별할 수 없는 경우 다음과 같이 나타낸다.

$$\begin{aligned} (x_i, x_j) &\in ind(P) \\ ind(P) &= \bigcap_{R \in P} ind(q) \end{aligned}$$

특히,  $P=Q$ 인 경우  $(x_i, x_j) \in ind(Q)$ 라면  $x_i$ 와  $x_j$ 는 정보 시스템에서 식별할 수 없는 것으로 한다.

$ind(q)$ 는 동치 관계이기 때문에 이것에 의한 X의 분할은 상집합이 된다.

$$X/ind(q) = \{[x_i] \mid x_i \in X\}$$

## 4. 모델의 응용 및 해석

### 4.1 문제 설정

어떤 병원에서 환자들이 진료를 위해 임상 검사를 실시한다고 하자. 그러면,  $x_i$ 는 환자,  $q_i$ 는 검사 항목,  $V_{q_i}$ 는 검사 결과라고 볼 수 있다. 그러면 다음과 같은 정보 시스템 테이블을 구성할 수 있다.

$x_1$ 부터  $x_5$ 까지의 5명의 환자가 각각  $q_1$ 부터  $q_4$ 까지의 4개 검사 항목에 대해 임상 검사를 실시했다. 이에 대한

[표 2] 환자들의 검사 항목에 대한 정보 시스템 테이블

환자(X)	검사항목(Q)			
	$q_1$	$q_2$	$q_3$	$q_4$
$x_1$	1	1	0	1
$x_2$	0	1	0	1
$x_3$	1	1	0	0
$x_4$	1	1	2	1
$x_5$	0	2	1	2

결과가 0부터 2까지의 3가지 형태로 얻어졌다. 여기서, 각 속성들에 대한 속성 값의 러프 집합의 유사성을 이용하여 객체들에 대한 분류를 수행할 수 있다.

### 4.2 러프 집합을 이용한 테이블의 해석

[표 2]를 러프 집합의 동치 관계 개념을 이용하여 분류해보자.

객체 집합 X를 다음과 같은 속성들로 분할한 결과는 다음과 같다.

$$\begin{aligned} X/ind(q_1) &= \{[x_1, x_3, x_4], [x_2, x_5]\} \\ X/ind(q_2) &= \{[x_1, x_2, x_3, x_4], [x_5]\} \\ X/ind(q_3) &= \{[x_1, x_2, x_3], [x_4], [x_5]\} \\ X/ind(q_4) &= \{[x_1, x_2, x_4], [x_3], [x_5]\} \\ X/ind(\{q_2, q_3\}) &= \{[x_1, x_2, x_3], [x_4], [x_5]\} \\ X/ind(Q) &= \{[x_1], [x_2], [x_3], [x_4], [x_5]\} \end{aligned}$$

앞의 분류에 의하면 분류의 기준이 되는 속성이 많을수록 더욱 세밀하게 분류된다. 즉, 속성  $q_2$ 에 의한 분류는  $[x_1, x_2, x_3, x_4]$ 와  $[x_5]$ 로 2 분류되지만 속성  $q_2$ 과  $q_3$ 에 의한 분류는  $[x_1, x_2, x_3], [x_4], [x_5]$ 로 3 분류된다. 이는 분류 기준이 강화될수록 객체 간의 유사성이 적어지기 때문이다.

[표 2]를 속성들의 집합 Q로 분류하면 다음 [표 3]을 얻을 수 있다.

[표 3] Q에 의해 분류된 테이블

$X/ind(Q)$	$q_1$	$q_2$	$q_3$	$q_4$
$\{x_1\}$	1	1	0	1
$\{x_2\}$	0	1	0	1
$\{x_3\}$	1	1	0	0
$\{x_4\}$	1	1	2	1
$\{x_5\}$	0	2	1	2

앞의 [표 3]에서 속성  $q_1$ 을 제거하면 다음 [표 4]를 얻을 수 있다.

[표 4] 속성  $q_1$ 이 제거된 테이블

$X/ind(Q)$	$q_2$	$q_3$	$q_4$
$\{x_1\}$	1	0	1
$\{x_2\}$	1	0	1
$\{x_3\}$	1	0	0
$\{x_4\}$	1	2	1
$\{x_5\}$	2	1	2

그런데, 속성  $q_1$ 을 제거하면 객체  $x_1$ 과  $x_2$ 가 식별 불가능해지기 때문에 속성  $q_1$ 을 제거할 수 없다. 그러면 이번에는 [표 3]에서 속성  $q_4$ 를 제거해보자. 그러면 다

음과 같은 표를 얻을 수 있다.

[표 5] 속성  $q_4$ 가 제거된 테이블

$X/\text{ind}(0)$	$q_1$	$q_2$	$q_3$
$\{x_1\}$	1	1	0
$\{x_2\}$	0	1	0
$\{x_3\}$	1	1	0
$\{x_4\}$	1	1	2
$\{x_5\}$	0	2	1

이 역시 객체  $x_1$ 과  $x_3$ 을 식별할 수 없기 때문에 속성  $q_4$ 를 제거할 수 없다. 그러나, 속성  $q_2$ 는 제거되더라도 다음 [표 6]과 같이 식별 불가능한 객체들이 없기 때문에 이는 생략 가능하다. 그러므로 속성 집합  $0$ 에 의해 분류된 결과는  $(q_1, q_3, q_4)$ 로 간략화시킬 수 있다.

[표 6] 속성  $q_2$ 가 제거된 테이블

$X/\text{ind}(0)$	$q_1$	$q_3$	$q_4$
$\{x_1\}$	1	0	1
$\{x_2\}$	0	0	1
$\{x_3\}$	1	0	0
$\{x_4\}$	1	2	1
$\{x_5\}$	0	1	2

결국, [표 2]로 주어진 테이블은 데이터베이스를 구성하는 객체의 식별을 위해서는 [표 6]으로 간략화시켜 검색할 수 있다는 장점이 있다.

## 5. 결론

정보 검색 시스템의 목표는 정확한 검색 결과의 생성이다. 그러나, 현실 세계에는 불명확한 질의에 대해서 가능성 한계 내에서 생성된 근사적 결과가 필요한 경우도 많이 있다. 이러한 요구에 대해 기존의 RDB는 만족스러운 결과를 생성하지 못해왔다. 그러나, 본 논문에서 제안하는 러프 집합을 이용한 관계데이터베이스 모델에서는 이러한 한계 공간 내에서의 검색을 수행할 수 있다는 강점을 보이고 있다. 본 논문에서는 단지 그 방법론만을 다루었지만 현실적인 응용 가능성은 매우 높다고 생각된다.

향후로는, 유사성에 의해 분류된 객체들에 러프 집합의 상한 근사와 하한 근사를 이용하여 의사 결정에 필요한 정보를 유도해내는 과정이 필요하며, 이와 함께 본 논문에서의 제안을 결합하여 실제 데이터베이스 모형으로 구현하는 작업이 필요하다. 또한, 퍼지 집합론의 가능성 속도와 러프 집합의 근사 공간 특징을 조합하여 불확실한 상황 하에서도 사용자의 기대 욕구를 충족시킬 수 있는 확장된 관계데이터베이스 모델에 대한 연구가 지속적으로 이루어져야 할 것이다.

## 6. 참고문헌

- [1] Z.Pawlak, Rough Sets, Theoretical Aspects of Reasoning about Data, Kluwer Academic Publisher, 1991.
- [2] F.E.Petry and T.Beaubouef, "A Rough Set Model for Relational Databases", LAQSF-GF-13, 1995.
- [3] Z.Pawlak, "Rough Sets", International Journal of Computer and Information Science, 1982.
- [4] M.Umano, "A fuzzy Database System", Information Process, 1982.

[5] F.E.Petry, "A Fuzzy Representation of Data for Relational Databases", Information Process, 1980.

[6] B.P.Buckles and F.E.Petry, "Uncertainty Models in Information and Database Systems", Journal of Information Science, 1985.