

주관투영도(主觀投影圖)를 활용한 실세계의 추상화

노태호*, 최인수**

Abstraction of the Real World Using Projective Maps

Tae-ho Roh*, In-soo Choi**

요약

주관투영도는 데이터베이스 설계자가 추상화 단계에서 사용자의 불완전한 관점에서 발견되지 못한 속성(covered attributes)을 찾아내어 사용자의 모델에 맞는 데이터베이스를 설계한다. 이런 데이터베이스는 고유한 의미를 유지하고 논리적으로도 실세계를 보다 정확히 표현하는 좋은 구조를 갖는다. 또한 주관투영도를 활용하여 설계된 데이터베이스는 사용자들에게 적은 수의 사용자 프로젝션(projection)으로 실세계를 기술할 수 있는 방법을 제시 한다.

Abstract

Projective Maps let designers find out uncovered attributes which are unexposed by users' incomplete view at abstraction step, and design proper database for users' model. This database keeps the original meaning and has a good structure to represent the real world precisely. Also, the database which is designed by using Projective Maps, presents a method which describes the real world by less projections to users.

I. 서론

데이터베이스와 데이터베이스 기술은 오늘날 컴퓨터의 보편화에 많은 영향을 주어 왔고, 특히 컴퓨터가 사용되는 거의 모든 분야에서 매우 중요한 역할을 담당 하고 있다고 생각한다.

과연 데이터베이스란 무엇인가? 데이터베이스는 서로 연관이 있는 데이터들의 집합이고 이 데이터들에는 알고 있는 사실(facts)에 대한 의미(meaning)를 내포하고 기록할 수 있는 특성이 있다[5]. 예를 들면, 이 논문을 구성하고 있는 단어들은 서로 연관된 데이터이므로 데이터베이스라고 할 수 있다. 그러나 데이터베이스에 대한 이러한 정의는 너무 일반적이기 때문에 데이터베이스를 다음과 같은 특성을 갖는 것으로 다시 정의하고자 한다.

- 1) 데이터베이스는 고유의 의미를 가지며 동시에 논리적으로 일치하는 데이터의 집합이다.
- 2) 의미 없이 데이터가 무작위로 분류된 것은 데이터베이스라고 하지 않는다.
- 3) 데이터베이스는 특별한 목적으로 설계, 구축되며 또한 운용된다.
- 4) 데이터베이스는 작은 세계라 불리는 실세계의 일부 분이다.[2]

따라서 데이터베이스의 정의를 감안해가면서 효율적으로 데이터베이스를 구축하기 위해서는 개발자들은 사용자들과 철저한 인터뷰를 통하여 사용자들의 활동영역을 잘 반영하는 사용자 모델(user's model)을 완전히 이해해야 한다. 이렇게 하기 위해서 개발자들은 데이터베이스 내에 저장된 정보(데이터)를 구별하고, 데이터의 구조와 데이터들 간의 관계를 정의한다. 즉, 데이터를 사용자의 관점(view)으로 추상화(abstraction)하여 모델링을 해야 한다. 여기서 모델이란 어떤 것을 구축하기 전에 이해할 목적으로 이를 추상화한 것이고, 추상화란 어떤 실체의 근본적이며 고유한 면에 초점을 맞추고 우연적인 성질을 무시하는 것이다. 즉, 개체(identities)들의 일시적인 특

성(properties)들은 무시하고, 개체의 상속 내용과 필수 요소에 중점을 두는 것을 뜻한다. 이는 시스템을 개발하기 전에 개체가 무엇(what it is)이며 무엇을 하는(what it does) 것인지에 중점을 두는 것을 뜻한다. 따라서 추상화를 한다는 것은 문제의 측면들을 선택적으로 검토하고, 목적상 중요한 측면을 분리 추출하고, 중요하지 않은 부분은 무시하므로, 어떤 상황에서 유일하고 정확한 모델은 존재하지 않고, 단지 적절한 모델이 존재하는가 아니면 부적절한 모델이 존재하는가에 대한 존재일 뿐이다. 그러므로 좋은 모델은 어떤 문제에 대해서 가장 중요한 부분을 포착하고 다른 부분은 생략한다고 정의 할 수 있다. 또한 추상화를 적절하게 하면 분석, 설계, 프로그램 구조, 데이터 구조 그리고 문서화작업 등에 똑같은 모델을 재사용할 수 있다[1]. 따라서 추상화는 데이터베이스 구축 시 정보를 모델링 하는 단계에서 가장 중요한 것이라고 할 수 있다. 그러나 분석 단계에서 사용하는 언어는 우리가 일상적으로 사용하는 자연 언어이기 때문에 많은 문제를 야기 시키는데, 이런 많은 문제들은 상세 설계 시 적절한 설계를 해주는 분석 과정에 대한 방법론의 부재로 야기되는 것이다[4]. 따라서 적절한 추상화를 한다는 것은 좋은 구조(well structure)의 데이터베이스를 설계하는 것이므로 앞에서 언급한 데이터베이스의 정의에 부합되며 사용자에게는 보다 정확한 실세계를 표현해 줄 수 있게 해주는 것이다.

본 연구에서는 데이터베이스를 기술하는데 사용된 데이터와 사용자의 불완전한 정보에 따른 완전하지 못한 사용자 관점에 대한 추론뿐만 아니라 사용자 관점들 사이의 관련 형식(pattern)을 정의하는 주관투영도(主觀投影圖)[3]를 활용하여 데이터베이스 설계자가 추상화 단계에서 사용자의 불완전한 관점에서 발견되지 못한 속성(covered attributes)을 찾아내어 사용자의 모델에 맞는 데이터베이스 즉, 고유한 의미를 유지하고 논리적으로도 실세계를 보다 정확히 표현하는 좋은 구조를 갖는 데이터베이스를 만들고자 한다. 또한 이런 좋은 구조의 데이터베이스로 사용자들에게 적은 수의 사용자 프로젝트(project)으로 실세계를 이해할 수 있는 방법을 제시하고자 한다.

II. 주관투영도(Projective Map)

이 절에서는 주관투영도의 특징과 적용 방법에 대하여 다룬다. 주관투영도는 추상화 시 프로젝트션(사용자 또는 설계자의 관점)으로 구성된 릴레이션(relation)을 입력으로 사용하여 안전하고 완전한 좋은 구조를 갖는 데이터베이스를 설계하고 또한 적은 수의 프로젝트션으로도 사용자(end user)에게 보다 더 명확하게 유니버스(universe)를 이해할 수 있게 해준다.

1. 주관투영도의 특징

주관투영도는 데이터베이스 설계자의 관점인 프로젝트션, 즉, 함수적 종속(functional dependency) [5,6]을 입력으로 사용하지만, 이 프로젝트션이 릴레이션 스키마에서 다른 스키마 보다 더 좋다는 형식적인 측정 방법은 없다.[3] 그리고 주관투영도를 구성하고 있는 각 셀(cells)들은 각각의 함수적 종속에 속하는 애트리뷰트(attributes)이다. 또한 주관투영도에 속하는 각 셀의 순서를 정하기 위하여 주관투영도의 입력인 프로젝트션을 진리표(truth table)를 이용하여 나열한다. 예를 들면 입력이 1개인 주관투영도와 진리표의 관계는 다음과 같다.

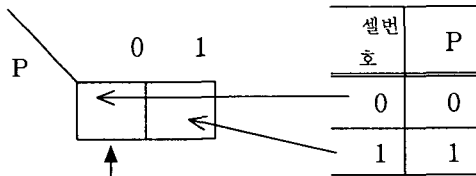
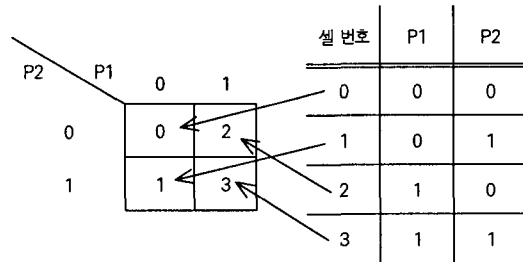


그림1. 입력이 1개인 주관투영도와 진리표 관계

그림 1.에서 주관투영도는 1개의 입력에 대해 2개의 값을 가지는 진리표에 대응하고, 서로 인접한 2개의 셀로 구성되어있다. 주관투영도의 셀 번호는 왼쪽부터 "0,1.." 순으로 정한다. 또한 맨 왼쪽 셀에 대한 P의 값은 "0"이고 그 옆에 있는 셀에 대한 P의 값은 "1"이다. 즉, 셀에 대한 P의 값이 "0" 혹은 "1"이란 뜻은 해당 셀의 애트리뷰트가 프로젝트션 P에 속하는 것인가 아닌가를 뜻한다.

다음은 입력이 2개인 주관투영도와 진리표의 관계를 나낸다.

그림2. 입력이 2개인 주관투영도와 진리표관계



다음 그림은 2 개의 입력을 갖는 또 다른 주관투영도의 예이다.

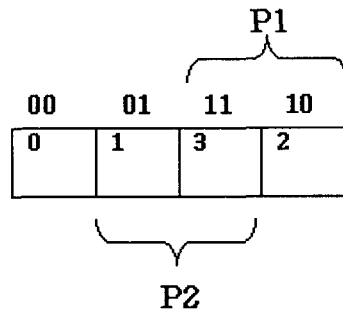


그림 3. 입력이 2개인 다른 표현의 주관투영도

그림 3 에서 셀의 순서가 0, 1, 2, 3이 아니라 0, 1, 3, 2 순서임을 알 수 가 있다. 이는 두 개의 입력인 P1 과 P2의 순서를 나열한 것과 셀의 순서가 서로 같다는 것을 나타내며 또한 두 개의 입력 P1과 P2는 인접한 셀에서 오직 한 입력만 다르다는 것을 나타내고 있다. 예를 들면, 셀 번호가 1과 3은 입력 P2는 같고(P2=1), 입력 P1은 셀 번호 1과 3에서 서로 다르다. 셀 번호 0과 2를 원통형으로 만들어 서로 인접한 경우도 위의 경우처럼 입력 P2는 같고 입력 P1은 서로 다르다. 그러므로 주관투영도에서 인접한 셀은 동시에 두 개 이상의 입력이 변할 수 는 없다는 것이다. 또한 주관투영도는 단지 입력의 있고(1:true), 없음(0:false)으로만 구성된다. 즉 어떤 관점에 대해서 그 관점으로 데이터를 사용하느냐 사용하지 않느냐이다. 즉, 데이터베이스 설계자의 관점인 함수적 종속에 속하지 않는 애트리뷰트는 필요 없다는 것이다. [그림 3]에서 P1, P2의 값을 각각 "0" 과 "1"로 나타내

있다. 여기서 "1"은 P1의 관점으로 보았다는 것이고, 0 "은 P1의 관점으로 보지 않다는 것을 뜻한다. P2의 경우도 P1의 경우와 마찬가지로이다. 따라서 서로 인접해 있는 각 셀의 관계를 살펴보면 대각선의 관계만 제외하면 수평적이든 수직적이든 서로 인접해 있는 셀은 주관투영도의 입력이 서로 반대 즉 보수(complemented) 관계에 있다는 것을 알 수 있다. 끝으로 주관투영도에서 한 가지 유의해야 할 사항은 셀 번호 0에 해당하는 애트리뷰트는 존재하지 않는다는 것이다.

2. 주관투영도 적용 방법

앞 절에서 주관투영도의 입력은 프로젝션이고 주관투영도의 각 셀들은 속성(attributes)라 정의 하였다. 다음과 같이 프로젝션과 이에 대응하는 속성이 있다고 하면 아래와 같이 표현할 수 있다.

표 1. 함수적 종속과 속성

P1	P2	Attributes
0	0	
0	1	A
1	0	
1	1	B

표 1 을 이용하여 주관투영도를 만들면 다음과 같다.

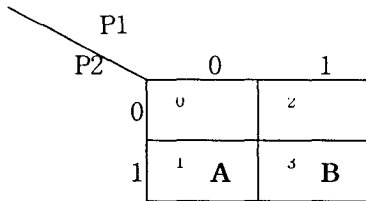


그림 4. 주관투영도

즉 속성A는 오직 프로젝션 P2의 관점에 속하고 속성 B는 프로젝션 P1, P2 의 모든 관점에 속한다. 다음 그림 5.는 그림 4.를 가지고 만든 또 다른 주관투영도이다.

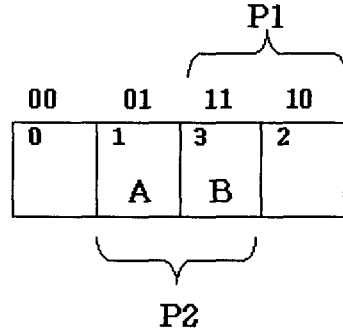


그림 5. 입력이 2개인 주관투영도

III. 프로젝션의 간략화

주관투영도에 있는 셀들은 주관투영도의 입력인 프로젝션에 대해 앞 절에서 언급한 것처럼 서로 보수 관계를 이용하여 주관투영도의 입력인 프로젝션들을 간략화 할 수 있는데 이것이 주관투영도의 또 다른 특징이다. 즉, 가로 혹은 세로로 인접한 셀들은 어떤 프로젝션에 대하여 서로 보수의 관계가 있으므로 서로 보수의 관계가 있는 이들 프로젝션들은 제거할 수 있다.[3]

1. 입력이 3개인 주관투영도의 간략화

다음 그림은 3개의 프로젝션에 대한 예제이다.

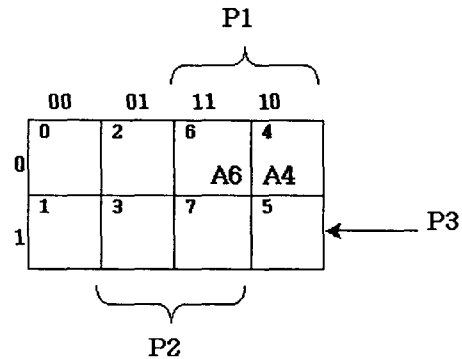


그림 6. 입력이 3 개인 주관투영도

그림 6에서 셀 번호 6 과 4 에 각각 해당하는 애트리뷰트가 있다고 가정한다면 이 애트리뷰트들은 입력 P1의 관점에 속하는 애트리뷰트들이고 P3의 관점에는 속하지 않는 애트리뷰트들이다.

또한 셀 번호 6 에 속하는 애트리뷰트(A6)는 입력 P2의 관점에는 속하고 셀 번호 4 에 속한 애트리뷰트(A4)는 입력 P2의 관점에는 속하지 않는 애트리뷰트 임에 틀림없다. 따라서 이를 수식으로 나타내면 아래와 같다.

$$A6 = P1 P2 \overline{P3} \quad \text{--- (1)}$$

$$A4 = P1 \overline{P2} \overline{P3} \quad \text{--- (2)}$$

위의 수식(1, 2)과 주관투영도의 특징은 서로 인접한 셀은 오직 하나의 입력이 서로 보수인 관계를 이용한다면 다음과 같다.

$$\begin{aligned} A6 + A4 &= P1 P2 \overline{P3} + P1 \overline{P2} \overline{P3} \\ &= P1 \overline{P3} \quad \text{--- (3)} \end{aligned}$$

식 (3)으로부터 우리는 여기서 프로젝션 1개가 줄었다는 것을 알 수 있다.

2. 입력이 4개인 주관투영도의 간략화

아래 그림은 주관투영도상에서 위치적으로 인접하지는 않지만 의미적으로 인접한 4개의 프로젝션에 대한 예제이다.

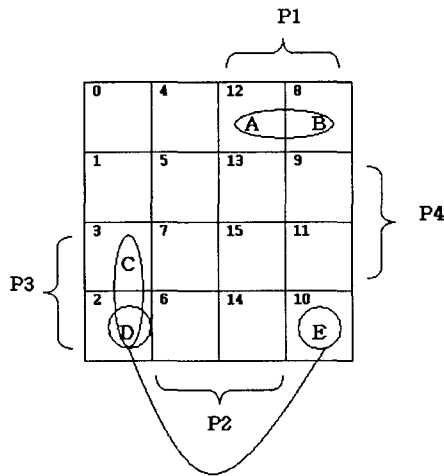


그림 7. 입력이 4 개인 주관투영도

그림7에서 왼쪽에 있는 4개의 셀(0,1,3,2)와 오른쪽에 있는 4개의 셀(8,9,11,10)들은 같은 행끼리 서로 인접하다. 그리고 맨 위쪽의 첫 번째 행에 있는 4개의 셀(0,4,12,8)과 맨 아래쪽 행에 있는 4개의 셀(2,6,14,10)들은 같은 열 끼리 서로 인접하다. 또한 그림7에 나타나 있듯이 서로 인접한 각 셀들은 묶여져 있다. 주관투영도에 있는 애트리뷰트를 수식으로 표현하면 다음과 같다.

$$A = P1 P2 \overline{P3} \overline{P4} \quad \text{--- (1)}$$

$$B = P1 \overline{P2} \overline{P3} \overline{P4} \quad \text{--- (2)}$$

따라서

$$\begin{aligned} A + B &= P1 P2 \overline{P3} \overline{P4} + P1 \overline{P2} \overline{P3} \overline{P4} \\ &= P1 \overline{P3} \overline{P4} \quad \text{--- (3)} \end{aligned}$$

이고

$$C = \overline{P1} \overline{P2} P3 P4 \quad \text{--- (4)}$$

$$D = \overline{P1} \overline{P2} P3 \overline{P4} \quad \text{--- (5) 이므로}$$

$$\begin{aligned} C + D &= \overline{P1} \overline{P2} P3 P4 + \overline{P1} \overline{P2} P3 \overline{P4} \\ &= \overline{P1} \overline{P2} P3 \quad \text{--- (6)} \end{aligned}$$

이다. 마지막으로 E의 경우는 D 혹은 B와 묶여질 수 있다. 여기서는 D와 묶여지는 경우만 취급한다.

$$E = P1 \overline{P2} \overline{P3} P4 \quad \text{--- (7)이고}$$

$$D = \overline{P1} \overline{P2} P3 \overline{P4} \quad \text{--- (5)이므로}$$

$$\begin{aligned} E + D &= P1 \overline{P2} \overline{P3} P4 + \overline{P1} \overline{P2} P3 \overline{P4} \\ &= \overline{P1} \overline{P2} P3 \quad \text{--- (8)} \end{aligned}$$

이상과 같이 식 (3), (6) 그리고 (8)에서 각각 1개의 프로젝션이 줄었음을 알 수 있다.

IV. 평가 및 검토

본 논문에서 제안한 주관투영도를 활용한 추상화방법에 대한 평가 및 분석을 다음과 같은 예제로 평가하였다. 어떤 데이터베이스 혹은 릴레이션에 다음과 같은 프로젝트션이 존재한다.

프로젝션 P1: SSN, P_number -> Hours

프로젝션 P2: SSN -> E_name, P_name

프로젝션 P3: P_number -> P_name 이다.

여기서 SSN은 주민 번호이고, P_number는 프로젝트 번호, E_name은 종업원 이름, Hours는 참여 시간이고 끝으로 P_name은 프로젝트 이름이다.

이를 바탕으로 주관투영도의 각셀에 들어갈 애트리뷰트의 순서를 아래의 표에 정의한다.[3]

표2. 애트리뷰트의 셀 번호

애트리뷰트	P1P2P3	g(P1,P2,P3)	Cell No
SSN	L L X	110	6
P_number	L X L	101	5
Hours	R X X	100	4
E_name	X R X	010	2
P_name	X R R	011	3

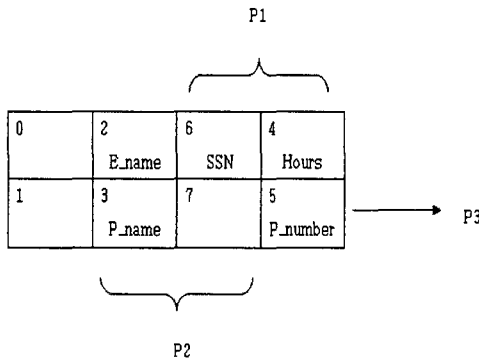


그림 8. 주관투영도에서의 각 애트리뷰트 위치

E_name과 P_name을 알기 위해서는 사용자는 P1 뿐만아니라 P2, P3 모두 사용해야한다. 하지만 앞 절의 방법을 사용하면 식(3)과 같이 2개의 프로젝트션을 줄 일 수 있다.

$$E_name = \overline{P1P2P3} \text{ --- (1)}$$

$$P_name = \overline{P1P2P3} \text{ --- (2) 이므로}$$

위 2 식을 합치면 식(3) 같다.

$$E_name + P_name = \overline{P1P2P3} + \overline{P1P2P3} = \overline{P1P2} \text{ --- (3)}$$

같은 방법으로 SSN과 Hours, SSN과 E_name, 그리고 Hours와 P_number를 r각각 합치면 1개의 프로젝트션이 줄었음을 알 수 있다. 또한 설계 시 발견하지 못했던 중요한 애트리뷰트를 발견하기 위하여 {3}같은 방법으로 확인한다.

표3. 주관투영도의 셀 중요도

프로젝션	셀 번호							
	0	1	2	3	4	5	6	7
P1					■	■	■	■
P2			■	■			■	■
P3		■		■		■		■
중요도	0	1	1	2	1	2	2	3

위 표3.에서 셀의 중요도를 확인하면 7번의셀이 가장 중요함을 알수 있고, 설계 시 발견되지 못한 어떤 애트리뷰가 이 셀에 포함되면 틀림없이좋은 구조를 갖을 것이다. 따라서 7번 셀에 애트리뷰트 "X"가 있다면 우리는 4개의 셀 (4,5,6,7)을 다음과 같이 묶을 수 있다.

$$SSN = \overline{P1P2P3} \text{ --- (1)}$$

$$Hours = \overline{P1P2P3} \text{ --- (2)}$$

$$X = \overline{P1P2P3} \text{ --- (3)}$$

$$P_number = \overline{P1P2P3} \text{ --- (4)}$$

따라서 위 식(1),(2),(3)과 (4)를 묶으면 아래와 같이

오직 1개의 프로젝션만으로도 고유한 의미를 그대로 유지하고 논리적으로도 실세계를 정확히 표현하는 좋은 구조를 갖는다.

$$\begin{aligned}
 &SSN + Hours + X + P - number \\
 &= P_1 P_2 \overline{P_3} + P_1 \overline{P_2} P_3 + P_1 P_2 P_3 + P_1 \overline{P_2} \overline{P_3} \\
 &= P_1 - - - (5)
 \end{aligned}$$

V. 결론

본 연구에서 주관투영도를 활용하여 데이터베이스를 기술하는데 사용된 정보와 사용자의 불완전한 프로젝션을 추론하여 사용자의 프로젝션들 사이의 관련 형식들을 정의하는 새로운 정량화된 추상화 기법을 제시하였다. 뿐만 아니라 본 연구에서 제안한 새로운 추상화 기법으로 설계된 데이터베이스는 사용자의 모델에 맞는 데이터베이스이기 때문에 고유한 의미를 유지하고 논리적으로도 실세계를 보다 정확히 표현하는 좋은 구조를 갖는 데이터베이스가 된다고 본다. 또한, 이러한 좋은 구조를 갖는 데이터베이스를 활용한다면 사용자들은 보다 적은 수의 사용자 프로젝션으로 실세계를 이해할 수 있게 된다고 본다. 비록 본 연구에서 제시한 주관투영도를 활용한 추상화 기법이 정량화되고, 데이터베이스 설계 시 하나의 중요한 도구로 활용되겠지만 여기에 대해서는 보다 많은 연구를 해야 한다고 생각한다.

참고문헌

[1] James Rumbaugh 외 Object-Oriented Modeling and Design, 7, Prentice Hall, 1991
 [2] Elmasri / Navathe, Fundamentals of Database Systems, The Benjamin/Cummings

Publishing Company, Inc, 1989

[3] Taeho Roh, Insoo Choi, The Multi-Projective Model: An Object-Oriented Logical Model, IJMS, 2001,4
 [4] Geoff Cutts, Structured Systems Analysis and Design Methodology, Second Edition, Blackwell Scientific Publication, 1991
 [5] David M. Kroenke Database Processing fundamentals, design & implementation seventh edition, Prentice Hall, 2001

저자 소개



노 태 호

승실대학교 산업·정보시스템
공학과 박사과정
Polytechnic University,
Computer Science MS.
동양공업전문대 산업체겸임
교수 역임
주요관심분야:
객체지향데이터베이스,
Data Mining, OLAP, MIS



최 인 수

서울대학교 산업공학 공학
박사 취득
영국 College of Librarianship,
Wales (Diploma), 日本 東
洋大學 経営學部 객원교수
현재 승실대학교 산업·정보시
스템공학과 교수에 재직 중
주요 관심분야 :
객체지향데이터베이스,
정보시스템 분석·설계,
ERP, MIS