퍼지 속성 공간을 이용한 객체 상호 작용에 관한 연구

1997 년 퍼지 및 지능 시스템 학회 추계 학술대회 제출 논문 (1997. 11. 17. Version)

이진호, 이전영
포항공과대학교 전자계산학과

(790-784) 경상북도 포항시 남구 효자동 산 31 번지
포항공과대학교 정보통신연구소 지능정보시스템연구실

TEL: 0562-279-{5662, 2242} FAX: 0562-279-5699
Email: {zino, jeon}@postech.ac.kr

A study on the Object Interoperability using the Fuzzy Property Space

Zino Lee, Jeonyoung Lee
IIS Lab./PIRL/CSE Dept./POSTECH

San 31, HyoJa Dong, NamGu, Pohang, 790-784, Korea Republic of

Abstract

퍼지 속성 공간(fuzzy property space)은 데이터베이스의 각 객체를 분류하고 분석하는데 유용한 도구로서 사용됨을 보였다[1]. 이는 수학적인 속성 집합 이론(property set theory)[2]에 근간을 두고 만들어진 이론으로, 데이터의 분석에 무척 유리한 도구로 사용될 수 있다.

본 논문에서는 근래에 들어 많은 연구가 이루어지고 있는 분산 데이터베이스 환경(distributed database management)에서 이를 응용해보고자 시도하였다. 즉, 분산 환경에서 어떠한 객체의 데이터를 상호 교환하고자 하는 간단한 상호 작용(object interoperability)을 수행함에 있어, 각 시스템은 이들 상호간의 규약에 의한 합치(object integration)를 이룰 수 있어야 한다. 여기에 퍼지 속성 공간을 이용하여, 가장 근사한 합치를 이룰 수 있도록 하는 것이다. 예를 들어, A 와 B 두개의 시스템에서 객체의 상호 작용을 수행한다 하면, A 시스템의 하나의 객체를 두개의 공통된 속성 공간에 위치시키고, B 라는 시스템에서 이를 다시 해석하여 자신의 데이터베이스에 입력으로 받아들이는 방식을 채택하여 상호 작용의 연산을 설계하는 방식이다.

I. 서론

데이터에 있어 객체의 상호 작용(object interoperability)에 관한 문제 해결에 관한 연구는 분산 환경의 다양한 정보 처리 시스템 환경(heterogeneous information system environment)에서, 상호 데이터를 교환하고, 이를 각각의 시스템에서 무리 없이 사용할 수 있도록 하는 이론적인 기반을 마련하는 것을 해결하고자 함이다. 기존의 연구 결과로는 OMG 의 CORBA 와 마이크로소프트社의 자사 표준인 OLE/DCOM, IBM 社의 OpenDoc 을 들 수 있으나, 이러한 표준안들은 각 응용 프로그램에 기반한 데이터의 상호 교환을 의미한 것일 뿐, 데이터베이스를 기반으로 하지 않고 있다. 본 논문에서는 데이터베이스를 기반으로 한 객체의 상호 작용에 있어, 하나의 데이터 객체가 다른 시스템 환경의 데이터베이스에 수입(importing)되었을 경우, 해당 객체와 가장 유사한 데이터 군에 포함되기 위한 이론적 기반을 마련하고자 함이다.

데이터베이스 시스템(DB system)은 다량의 데이터를 무결성(consistency)을 유지하면서, 사용자의 요구에 따라 데이터의 입력과 삭제는 물론, 적절한 질의(user query)에 따라, 요구 조건에 맞는 데이터를 빠짐없이 추려내는 컴퓨터 시스템을 지칭한다. 이러한 데이터베이스 시스템을 구현하기 위해서는 우선 실세계의 정보를 자료 형식을 빌어 기술하는 개념적인 도구인 데이터 모델(data model)을 중심으로 이론적인 기반을 마련한다[1]. 이 데이터 모델은 본질적으로 실세계에 존재하는 데이터들을 무결성을 유지하면서(consistency constraints), 데이터 자체와 그 구조(structural data model)를 표현한 것들이다. 이들 수학적인 데이터 모델들의 기반이 되는 본질은 개체(object)와 그 속성(attribute), 그리고 그 속성 값(attribute value)으로 요약된다.

본 논문에서 접근하고자 하는 방식은 기존의 속성 집합 모델에 퍼지 집합(fuzzy set)[2] 이론을 도입한 퍼지 속성 집합(fuzzy property set)을 이용한 다양한 환경의 데이터베이스 상호간 객체 상호 작용(object interoperability in the heterogeneous database environment)에 관한 연구이다. 따라서, 본 논문의 구성은 우선 이러한 기반 지식을 이루는 퍼지 속성 집합 모델과, 퍼지 속성 공간(fuzzy property space)에 대한 소개하고, 이를 대상으로 각 객체가 다른 환경의 데이터베이스로 수입되는 경우, 이렇게 환경에 적용될 수 있는가 여부를 그 이론적인 기반에 따라 설명할 것이다.

II. 속성 집합 모델

이것은 실세계의 데이터를 수학적인 모델로 표현하는 방식으로 수학에서 논하는 집합(set)과 관계(relation), 그리고 함수(function)로 구성이 된다[3].

기본 구성

- 실세계에는 개체(object)의 집합인 Θ가 있다.
- 그리고, 그 개체를 표현하는 성격(attribute)들의 집합 AT 가 있으며,
- 각 성격이 가질 수 있는 값(value)들의 집합 VAL 이 있다.

$$VAL = \cup_{a \in AT} VAL_a$$

- 그리고, 여기에서 정보(information)란 각각의 개체가 어떠한 성격에 어떠한 값을 가지고 있는가 하는 함수(function)로 표현이 되며, 이를 정보 함수(information function) I_f라 한다.

$$I_f : \Theta \times AT \rightarrow VAL$$

이러한 기본 구성을 바탕으로 다음과 같이 속성 집합 모델이 정의된다.

정의 1 속성 집합 모델 (property set model)

- 실세계 개체 집합의 부분 집합으로 개체 집합 Θ가 정의된다.

$$\Theta = \{ \theta_1, \theta_2, \ldots, \theta_m \}$$

- 어떠한 개체(object)가 가질 수 있는 속성(property)들의 집합을 정의한다.

$$\Pi = \{\pi_1, \pi_2, \ldots, \pi_n\}$$

- 여기에서 속성이란, 성격-값의 관계(relation)로 정의된다.

$$\pi = (a, \upsilon), a \in AT, \upsilon \in VAL$$

- 개체(object)와 속성(property) 사이에는 연관된 관계(relation) 집합인 has-property 를 정의할 수 있다.

$${}_{\theta}P_{\pi} = (\theta, \pi) \in P$$

이렇게 정의된 속성 집합 모델을 기반으로, 개체(object)는 속성들의 집합으로 정의될 수 있으며, 속성(property)은 다시 개체들의 집합으로 정의될 수 있다. 즉, 다음과 같이 정의할 수 있을 것이다.

보조정리 1

속성 집합 모델에서, 속성은 해당하는 속성을 가진 개체들의 집합으로 정의되고, 개체는 자신이 가진 속성들의 집합으로 정의된다.

$$[\pi] = \{\theta | \theta \in \Theta, (\theta, \pi) \in P\}$$
$$[\theta] = \{\pi | \pi \in \Pi, (\theta, \pi) \in P\}$$

속성 집합 모델에서는 속성들로서 정의되는 개념 공간(conceptual space)을 정의할 수 있다. 속성 공간은 각 속성들이 하나의 축을 정의하는데, 이렇게 할 경우, 각각의 개체들은 고유의 위치를 지정 받게 된다.

정의 2 단위 개념(atomic concept)

속성들로서 정의되는 개념 공간(conceptual space)에서, 각각의 고유 위치를 정의하는 방식으로 특정한 속성들 혹은 속성의 여집합들의 교집합으로 정의된다.

$$[c_0] = [\pi_1]^c \cap [\pi_2]^c \cap \ldots \cap [\pi_n]^c,\ [\pi]^c = \Pi - [\pi]$$
$$[c_1] = [\pi_1] \cap [\pi_2]^c \cap \ldots \cap [\pi_n]^c$$
$$\ldots\ldots$$
$$[c_n] = [\pi_1] \cap [\pi_2] \cap \ldots \cap [\pi_n]$$

여기에서 정의되는 단위 개념은 개념 공간(conceptual space)이 정의됨에 따라서 정의되는 것이며, 개념 공간은 몇몇 개의 속성이 정의됨에 따라 정의되는 것이다. 또한, 정의되는 단위 개념의 개수는 개념 공간을 정의하는 속성들의 멱집합(power set) 갯 수와 같다. 즉, 다음과 같은 식이 성립된다.

$$C = \{c_1, c_2, \ldots, c_N\}$$
$$N = 2^n - 1$$
$$n = card(\Pi),\ card(B) = \text{cardinality of set B}$$

개념 공간에서 정의되는 각각의 개체들의 위치는 다음과 같다.

보조정리 2

개념 공간 안에서 특정한 개체 θ는 다음과 같이 정의되는 단위 개념(atomic concept)에 위치하게 된다.

$$c_\theta = \left(\bigcap_{\pi_i \in [\theta]} [\pi_i] \right) \cap \left(\bigcap_{\pi_j \notin [\theta]} [\pi_j]^c \right)$$

데이터베이스에서 이루는 객체 모델은 그 형태가 유사한 객체들의 집합을 클래스(class)로 묶어 관리하는데, 클래스는 시스템에 따라, 여러 가지 방식으로 정의될 수 있다. 하지만, 일반적으로 클래스란 같은 성격(attribute)들을 가지는 객체들의 집합이라 정의한다면[1], 속성 공간은 다음과 같이 정의될 수 있다.

정의 3 속성 공간 (property space)[5]

속성들로서 정의되는 개념 공간(conceptual space)들의 총 합집합으로, 각 성격들의 곱으로 정의된다. 이 공간 안에는 각 객체들의 고유 위치를 정의할 수 있다.

$$PS_{CID} = \times_i AT_i$$

여기에서, i는 각 클래스에 정의된 성격(attribute)의 갯수이며, AT_i는 각 클래스별로 정의된 성격을 의미한다.

이들 속성 공간(property space)은 다시, 존재할 수 있는 모든 단위 개념의 합집합으로 정의될 수 있다. 즉, 각각의 성격(attribute)들은 그 값과 함께 속성(property)으로 정의되고, 이들을 공간적인 개념으로 표현한 것이 다시 속성 집합이므로, 속성 공간은 모든

[1] 여기에, 상속성(inheritance) 혹은 복합성(polymorphism) 등의 다양한 클래스 성격에 관한 정의는 [5]를 참조하기 바란다.

단위 개념의 합집합으로 정의될 수 있다.

보조정리 3

속성 공간(property space)은 단위 개념(atomic concept)의 합집합으로도 정의가 된다.

$$PS_{\mathrm{CID}} = \cup_i c_i$$

여기에서, i는 각 클래스에서 정의될 수 있는 모든 속성(property)의 갯 수(n = card(Π))가 생성해내는 단위 개념만큼(2^n-1) 존재한다.

III. 퍼지 속성 집합 모델

퍼지 속성 집합은 앞에서 정의한 속성 집합 모델에 퍼지 집합 이론(fuzzy set theory)을 도입하여 확장한 것이다. 퍼지 이론을 도입하는 방안으로 우선 정의된 개념 공간의 구분이 모호해지도록 정의하는 방안을 고려할 수 있을 것이다.

따라서, 각 단위 개념(atomic concept)은 퍼지한 속성의 정의에 있어, 특정한 멤버 값(β)을 가지고 참여하게 된다. 즉, 어떠한 퍼지한 성질을 가진 속성이 퍼지 집합(fuzzy set)으로 정의되고, 이름이 B라 가정한다면, 어떤 단위 개념 c는 멤버 값 β로서 B의 원소로 정의될 수 있다.

$$(\beta, c) \in B, \text{ where } \mu_B(c) = \beta$$

따라서, 각각의 단위 개념들은 퍼지한 성격을 가지지 않지만, 개념 공간은 퍼지한 성격을 가진 속성들의 집합으로 정의될 수 있다. 퍼지 속성 집합 모델에서 각 속성들은 이러한 멤버 값을 가진 단위 개념들의 합집합으로 정의된다.

$$[\pi] = \{(\beta_i, c_i) \mid 0 \le \beta_i \le 1, c_i \in C\}$$

정의된 개념 공간에서 각 개체들은 어떻게 정의가 될까? 개념 공간의 속성상 개체는 특정한 단위 개념에 위치하게 된다. 다만, 각각의 단위 개념이 모여 퍼지한 속성을 정의하게 되고, 각 개체는 이렇게 정의된 퍼지한 속성들 혹은 속성의 여집합들의 교집합으로 정의될 것이다.

따라서, 특정한 퍼지 속성이 어떠한 개체를 정의함에 있어 어느 정도의 신뢰도(degree of confidence)를 갖는가 하는 척도가 필요하게 되며, 이를 α라 정의한다.

정의 4 개체 정의에 있어 퍼지 속성의 신뢰도

개체는 고유한 단위 개념(atomic concept)의 위치를 지정받는다. 퍼지 속성 집합 모델에서 단위 개념은 퍼지 속성들 혹은 속성의 여집합들의 교집합으로 정의되며, 이때 특정 속성이 어떠한 개체를 정의함에 있어 사용되는 신뢰도는 다음과 같이 정의된다.

$$(\alpha, [\pi]) = (\alpha, \{(\beta_i, c_i) \mid 0 \le \beta_i \le 1, c_i \in C\})$$
$$= \{(\delta_i, c_i) \mid \delta_i = 1 - \alpha + \beta_i(2\alpha - 1), c_i \in C\}$$

여기에서 만약 α가 1이라면, [π]의 속성을 완전히 인정하는 것이고, α가 1이라면 [π]의 여집합 속성을 인정하게 되는 것이다.

여기에서 각각의 개체는 정의에 따라 다음과 같이 퍼지한 속성들의 교집합으로 정의된다.

보조정리 4

개체(object) θ는 각 신뢰도를 가진 속성들의 교집합으로 정의된다. 이 교집합은 퍼지 개념 공간(fuzzy conceptual space)에서 퍼지 단위 개념(fuzzy atomic concept)을 정의하는 것이다.

$$\theta = \bigcap_{i=1}^{n} \{(\alpha_i, [\pi_i])\}$$

위의 보조 정리 3과 정의 3에 따라서, 각각의 개체는 단위 개념(atomic concept)의 교집합으로 재정의 될 수 있고, 이 때, 새로이 정의되는 δ_{ij}는 다음과 같이 정의된다.

$$\delta_{ij} = 1 - \alpha_i + \beta_j(2\alpha_i - 1)$$

따라서, 퍼지 속성 집합 모델에 있어, 개체는 다음과 같이 정의된다.

$$\theta = \bigcap_{i=1}^{n} \left\{ \left(\alpha_i, \bigcup_{j=0}^{N} \{(\beta_j, c_j)\} \right) \right\} = \bigcap_{i=1}^{n} \left(\bigcup_{j=0}^{N} \{(\delta_{ij}, c_j)\} \right)$$

그런데, 퍼지 집합 이론에서 교집합을 수행함에 있어 그 멤버 값은 항상 작은 쪽(minimum)을 택하게 되어 있으므로, 결국 퍼지 속성 집합 모델에서 개체의 정의는 다음과 같이 정의된다.

보조정리 5

퍼지 속성 집합 모델에서 개체(object) θ는 다음과 같이 단위 개념(atomic concept)의 합집합으로 정의된다.

$$\theta = \bigcup_{j=0}^{N} \left(\bigcap_{i=1}^{n} \{(\delta_{ij}, c_j)\} \right) = \bigcup_{j=0}^{N} \{(min\{\delta_{ij}\}, c_j)\}$$

여기에, 상호 유사한 성격들을 가진 객체들의 모임을 클래스(class)라 정의하였고, 클래스에 따른 속성 공간(property space)의 정의를 언급한 바 있다. 따라서, 퍼지 속성 공간은 각 단위 개념이 퍼지 집합 B에 속할 가능성을 고려하여, 다음과 같이 정의될 수 있다.

보조정리 6

퍼지 속성 공간 모델에서는 각 퍼지 단위 개념의 합집합으로 퍼지 속성 공간이 정의된다.

$$PS_{FuzzyCID} = \bigcup_{i=0}^{N} \{(\beta_i, c_i)\}$$

IV. 데이터 분석 알고리즘 개요

퍼지 속성 공간 모델(Fuzzy Property Space Model)은 데이터베이스 시스템에서 각 데이터들을 공간 내 특정 위치에 배치함으로써, 데이터 상호간의 관계를 계산할 수 있다. 이러한 성향은 새로운 데이터가 수입(importing)됨에 있어, 가장 유사한 위치가 어디일지 계산하는 곳에도 활용될 수 있음을 의미한다.

간단하게 설명하자면, 하나의 데이터가 입력됨에 있어, 해당되는 적절한 퍼지 속성 공간내의 위치를 파악한 뒤, 가장 적용이 가능한 근사한 형태로 변환하여, 활용하는 방안을 이론적으로 뒷받침하고자 함이다. 이러한 계산을 하귀 위한 간단한 알고리즘의 설명은 다음과 같다.

단계 1 현존하는 데이터베이스에 적절한 클래스를

지정하면, 해당 클래스에 맞는 퍼지 속성 공간을 계산한다. 이 속성 공간은 각 클래스별로 다를 수 있다.

단계 2 각 객체를 이 퍼지 속성 공간에 위치하여, 특별한 개체 군을 다시 추려낸다. 이렇게 함으로써, 어떠한 개체 군이 존재하며, 수입된 하나의 개체가 어느 개체 군에 유사한지를 판별하는데 응용하도록 한다.

단계 3 수입된 개체와 가장 유사한 개체 군을 찾아낸다.

단계 4 이제 수입된 개체의 데이터를 기반으로, 현재의 데이터베이스 시스템에 적용이 가능한 형태의 변환을 수행한다.

이 때, 속성 공간을 이루는 속성들은 흔히 질의 과정에서 프로젝션(Projection)에 해당하는 속성들을 지칭한다. 즉, 일련의 데이터들을 일정한 속성을 기준으로 분석하기 위해서는 데이터들을 원하는 속성만을 기준으로 추려내는 조건이 필요하고, 이 때 질의 결과 데이터는 프로젝션된 속성들이 형성하는 개념 공간 내에 위치하게 된다. 따라서, 각각의 개체가 가지고 있는 속성들의 집합은 모두 제각기 다를지 모르겠지만, 질의 결과 정리된 데이터들은 모두 같은 속성을 기준으로 표현될 것이다.

V. 데이터 분석 모델

임의의 두개 퍼지 집합의 유클리드 거리(Euclidean Distance)는 각 퍼지 집합을 이루고 있는 원소들의 멤버 값을 대상으로 거리를 측정한다.

보조정리 5 에 따라서 각 개체(object)들은 단위 개념과 어떠한 퍼지 집합의 멤버 값의 관계(relation)를 원소로 가지는 퍼지 집합으로 정의가 된다. 따라서, 임의의 두개 개체의 유클리드 거리는 다음과 같이 정의될 수 있다.

$$\delta(\theta,\theta')=\sqrt{\left(\sum_{i=0}^{N}\left(\delta_i-\delta_i'\right)^2\right)}$$

그런데, 거리 측정에 필요한 δ_i는 어떠한 단위 개념(atomic concept)이 특정한 개체를 퍼지 개념 공간 내에서 정의함에 있어 다시 정의되는 멤버 값이므로, 이는 다시 다음과 같이 표현될 수 있다.

$$\delta(\theta,\theta')=\sqrt{\left(\sum_{i=0}^{N}\left(\mu_\theta(c_i)-\mu_\theta(c_i')\right)^2\right)}$$

여기에서 유클리드 거리가 가까우면 가까울수록 두개의 개체는 같은 성향을 나타낸다. 퍼지 집합에서 유클리드 거리를 이용하여 유사도(similarity)를 다음과 같이 정의할 수 있다.

$$\sigma(\theta,\theta')=1-\sqrt{\frac{\left(\sum_{i=0}^{N}\left(\mu_\theta(c_i)-\mu_{\theta'}(c_i')\right)^2\right)}{N}}$$

모호성을 표현하는 수학적인 모델로서는 퍼지 집합 이론이 이용되지만, 모호 데이터 분석을 위해서는 러프 집합 이론(Rough Set Theory)[4]이 편리하다.

러프 집합을 정의함에 있어서는 두 개체의 유사도가 중요한 역할을 하고, 임의의 두개의 개체의 유사도가 τ보다는 크다고 하고, 이들의 관계 집합을 $^{\tau}C$ 라 정의한다면, $^{\tau}C$ 는 다음과 같이 정의된다.

정의 4 유사도가 τ보다 큰 개체들의 관계 집합

$$(\theta,\theta')\in{}^{\tau}C \Leftrightarrow \sigma(\theta,\theta')\geq\tau$$

이 관계 집합을 이용하여 러프 집합을 정의할 수 있다. 전체 데이터를 대상으로 분석을 수행하므로, 그 대상은 각각의 개체가 아닌, 개체들의 멱집합(power set)이 될 것이다. 따라서, 러프 집합은 개체들의 멱집합을 대상으로 정의되며, 그 내용은 다음과 같다.

$$T\in 2^{\Theta}$$

$$\mathrm{low}(T)=\{\theta\mid(\theta\in T)\wedge\forall(\theta,\theta')\in{}^{\tau}C\wedge(\theta'\in T)\}$$

$$\mathrm{upp}(T)=\{\theta\mid(\theta\in T)\vee((\theta,\theta')\in{}^{\tau}C\wedge(\exists\theta'\in T)\}$$

$$\mathrm{bnd}(T)=\mathrm{upp}(T)-\mathrm{low}(T)$$

VI. 결론 및 향후 연구 방향

본 논문에서는 개념적인 퍼지 속성 공간 모델을 이용하여, 데이터베이스의 데이터들이 여러가지 환경에서 상호 작용할 수 있는 모델을 제안하였다. 분석을 위한 데이터를 위해 유클리드 거리를 측정하는 방법과 러프 집합을 정의하는 방법을 이용하였다.

하지만, 본 시스템을 구현하고 응용함에 있어서는 그 계산 속도가 문제가 된다. 따라서, 보다 속도 개선을 위한 연구가 있어야 할 것이다.

VII. 참고 문헌

[1] 임정훈, 이진호, 이전영, "객체-집합을 기반으로 하는 데이터 모델의 구조적인 정의", 1995 년 한국정보과학회 추계학술대회 발표 논문집, pp. 105~108 1995

[2] Lotfi A. Zadeh, "Fuzzy Sets", Inf. Control 8, pp. 338~353, 1965

[3] Michael Hadjmichael, S.K. Michael Wong, "The Fuzzy Property Set Model: A Fuzzy Knowledge Representation for Inductive learning", Proc. Of the 3rd IEEE Conference on Fuzzy System, pp. 684~689, 1994

[4] Z. Pawlak, "Rough Sets", International Journal of Information and Computer Science, 11, pp. 344-356, 1982

[5] Zino Lee, Jeonyoung Lee, "An Object-Set based Data Model : The Structural Definitions with Operational Rules", Intl. Conf. on Computer and Industrial Engineering (ICC&IE 96), Oct. 1996, KyoungJu, Korea.