

양기준 선호모형: 비 정형적 자료검색을 위한 의사결정 모형

이 춘 열*

Double Anchors Preference Model (DAPM) : A Decision Model for Non-binary Data Retrieval

This paper proposes a new referential model for data retrieval as an alternative to exact matching. While exact matching is an effective data retrieval model, it is based on fairly strict assumptions and limits our capabilities in data retrieval. This study redefines data retrieval to include non-binary data retrieval in addition to binary data retrieval, proposes Double Anchor Preference Model(DAPM), and analyzes its logical characteristics. DAPM supports non-binary data retrieval. Further, it produces the same result as exact matching for the conventional binary data retrieval. These findings show that, at the logical level, the proposed DAPM retains all the desirable features for data retrieval.

I. 서 론

1960년대 처음 소개된 이후, 대부분의 데이터베이스 시스템은 아래와 같은 이분적 검색기준을 가정하고 있다.

“김포 국제공항에서 50km 이내에 위치하는 호텔들?”
또는

“홍 길동과 같은 부서에 근무하며, 홍 길동보다 급여가 많은 종업원의 이름은?”

이러한 전통적 검색기준들은 논리적 명제로의 전환이 가능하며, 이들 명제에 대하여 데이터베이스

내 모든 저장자료들은 진 또는 위 의 이분적 값을 지닌다. 일 예로서, 상기 호텔 검색기준에 대하여, 김포 국제공항으로 부터 40km에 위치한 호텔_A는 진 의 값을 가지며, 90km에 위치한 호텔_B는 위 의 값을 가진다.

이러한 논리적 근거에 기초하여, 전통적 자료검색은 이분논리 (Boolean Logic)에 기초하여 발전하여 왔다. 이분논리는 논리적 단순성, 명료성 등의 장점을 지닌다. 그러나, 이에 대한 반대급부로, 일반적 자료검색 시 경험하는 다양한 검색요구를 원활히 충족시킬 수 없다는 제한성을 지닌다 [Lipski,

* 한국통신소프트웨어연구소

1979; Wong, 1982].

일반적 자료검색의 예로서, 공장부지 선정을 위한 데이터베이스 검색을 고려하여 보자. 부지 선정을 위한 검색기준은 아래와 같다고 가정하자.¹⁾

- 공업용수 기준: 최소한 하루 평균 2000갤론의 공업용수 공급이 가능하여야 하며, 예상치 않은 초과수요에 대비하기 위하여 하루 최대 6000갤론 이상의 공업용수 공급이 가능하면 이상적이다.

- 위치 기준: 원자재의 특성상 공장 부지는 생산지로 부터 40km 이내에 존재하여야 하며, 가까이 위치할 수록 바람직하다.

본 공장부지의 선정 예는, 데이터베이스내 모든 가용부지를 검색기준을 만족시키는 부지와 그렇지 않은 부지로 양분할 수 없다는 점에서, 기존의 자료검색 예와 비교하여 근본적 차이점을 지닌다. 구체적으로, 선정기준의 하나인 공업용수를 고려하여 보자. 이 경우, 일일 공급량 2500 갤론은 공업용수 기준을 완전히 충족시킨다고 볼 수 없으며, 그렇다고 충족시키지 않는 것으로 완전히 제외시킬 수도 없다. 또한 일일 공급량 2500 갤론과 2700 갤론은 선정기준을 만족시킴에 있어 차이가 있다. 이상의 논리는 위치 기준에 대하여서도 공히 적용된다.

상기 예를 통하여 개괄적으로 예시된 바와 같이, 기존의 이분 논리에 기초한 자료검색모형은 다음과 같은 제한성을 지닌다:

- (1) 비 이분적인 검색질의의 표현 (query formulation)이 어려우며,
- (2) 비 이분적인 검색질의의 효율적 해석 (query interpretation)이 용이하지 않다.

본 논문은, 이분 논리에 기초한 자료검색모형이 지니는 제한성을 해결하기 위한 대안으로서 비 이분적 자료검색모형을 제시한다. 이를 위하여, 자료검색을 이분적 검색기준뿐만 아니라 비 이분적 검

색기준을 포함하도록 확대하며, 확대 정의된 자료검색의 효과적 지원을 위하여 양기준 선호모형 (DAPM: Double Anchor Preference Model)을 제시한다.

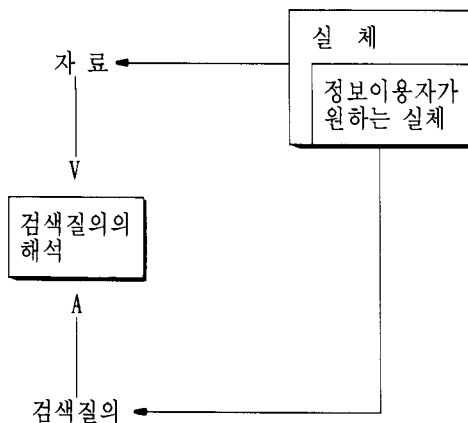
이하 II장은 의사결정의 관점에서 자료검색을 재정의하며, 이에 기초하여 자료검색을 정형적 자료검색과 비 정형적 자료검색으로 분류, 정의한다. III장에서는 비 정형적 자료검색을 위한 의사결정모형으로서 양기준 선호모형 (DAPM: Double Anchor Preference Model)을 제시한다. IV장은 제시된 양기준 선호모형의 특성을 분석, 평가하며, 끝으로 V장은 본 연구의 주요 결론 및 추후 연관 과제를 요약 제시한다.

II. 자료검색의 분류: 정형적 자료검색과 비 정형적 자료검색

1. 자료검색의 개념적 흐름도

일반적으로 자료검색은 데이터베이스내에 저장된 자료 (data)를 이용하여 정보 이용자가 원하는 실체 (relevant real world entities)를 검색하는 과정으로 정의된다 [Lyytinen, 1987]. 여기서 실체란 실 현상에 존재하는 물리적 또는 개념적 개체 (예: 공장부지, 학생, 거래, ...)이다. 이러한 개체의 특성은 데이터베이스내에 자료로 저장되며, 저장된 자료를 대상 실체의 속성이라 한다. 부연하면 데이터베이스내 실체는 저장 속성의 집합이며, 이는 물리적 단위인 실 현상의 개체에 대응하는 개념적 단위이다. 이상의 기본개념에 기초하여 본 논문은 자료검색을 [그림 1]과 같이 개념화한다.

1) 가용한 공장부지, 이들의 위치 및 공업용수 공급량은 데이터베이스에 저장되어 있다고 가정한다.



[그림 1] 자료검색의 개념적 흐름도

개념적 흐름도에 제시된 바와 같이, 자료검색은 자료와 검색질의의 대응 (matching) 과정으로 단순화된다. 여기서 자료는 실 현상의 실체 및 그들의 속성의 기록이며, 검색질의는 정보이용자가 찾고자 하는 실체 및 속성을 기술한 것이다. 따라서 자료검색은 이들 자료와 검색 질의가 실 현상의 실체 및 정보이용자가 찾고자 하는 실체의 속성을 각각 어떻게 대변하느냐에 따라 다른 특성을 지닌다. 요약하면, 자료검색은 자료와 검색질의의 속성에 따라 상이한 특성을 지닌다.

2. 자료 특성 분류

본절은 자료검색의 분류 및 특성의 정의를 위한 첫 단계로서 자료의 특성을 정의 분류한다. 자료의 특성 분류로서는, 데이터베이스에 저장된 자료가

실 현상에 존재하는 실체의 속성을 어떻게 대변하느냐에 따라, 완전자료(complete data) 검색, 부분자료(partial data) 검색, 누락속성(missing attributes) 검색 등으로 분류된다.

완전자료란 실 현상의 모든 실체 (예: 학생)에 대하여, (1) 정보이용자가 요구하는 속성 (예: 키)이 데이터베이스에 저장되어 있으며, (2) 각 실체의 속성은 상호 독립적인 단일 값 (crisp values; 예: 170cm)을 지님을 의미한다. 부분자료와 누락속성은 완전자료의 두 조건이 만족되지 않는 경우이다. 부분자료란 어떠한 실체 (예: 학생_A)의 속성 값이 상호독립적인 단일 값이 아닌 범위 값²⁾ (예: 170cm 부터 180cm 사이)을 지니는 경우이며, 누락속성이란 모든 실체에 대하여 정보이용자가 원하는 속성의 값이 저장되어 있지 않은 경우 (예: 키에 대한 자료가 데이터베이스에 저장되어 있지 않음)이다.

3. 검색질의 특성 분류

검색질의는 정보 이용자가 대상 실체에 대하여 어떠한 형태의 선호도를 지니느냐에 따라 다른 특성을 지닌다. 선호도란 정보이용자가 대상 실체를 선호하는 정도이다. 자료검색시 정보이용자의 선호도는 대상 실체의 선형배열³⁾ (linear ordering) 로 표현된다. 따라서 자료검색질의는 정보이용자가 대상 실체들에 대하여 총체적으로 몇개의 선호 계급을 가지느냐에 따라 서로 상이한 특성을 지닌다.

자료검색에 있어 최소 선호계급의 수는 둘이다.

- 2) 범위 값은 퍼지이론에서 제시된 퍼지 값과 개념상 동일하다. 범위 값의 가장 극단적인 경우는 null_value이다. null_value는 속성값 영역 (attribute value domain)내의 모든 값을 지닐 수 있음을 의미한다 [ANSI/X3/SPARC, 1975].
- 3) 선호도는 의사결정자가 대상실체에 대하여 지니는 총체적 선호 계급의 수와 선호계급간의 선호정도의 차이에 의하여 결정된다. 일반적 선호모형에 있어 선호정도의 차이는 절대값으로 표시된다. 그러나 자료검색에 있어 선호계급간의 선호정도는 절대값보다 상대값이 보다 큰 의미를 지닌다. 즉 어느 실체가 먼저 또는 늦게 검색되어야 하는가의 상대성이 의미를 지니며, 이러한 상대성은 실체의 선형배열로 표현된다.

그러나 최대 선호계급의 수는 대상 실체의 수에 따라 무한히 증가할 수 있다. 따라서 검색질의는 정보이용자가 지나는 선호계급의 수 (preference cardinality)에 따라 무한히 분류 정의될 수 있다. 본 논문은 특성 정의의 편의상 선호계급의 수가 둘인 경우와 둘이상인 경우로 구분한다.

선호계급의 수가 둘인 경우를 이분적 선호도를 지닌다고 한다. 이분적 선호도는 모든 가능 실체들이 이단계 (예: '좋아함'과 '싫어함')의 상호 독립된 집합으로 구분되는 선호도를 말한다. 일 예로, "홍길동과 같은 부서에 근무하는 종업원"의 검색에서, 정보이용자의 선호도는 대상 종업원이 홍길동과 같은 부서에 근무하는 경우와 그렇지 않은 경우에 대하여 서로 다르게 양분된다.

선호계급의 수가 둘이상인 경우를 비이분적 선호도를 지닌다고 한다. 비이분적 선호도는 가능한 대상 실체들이 삼단계 (예: '좋아함', '보통', '싫어함') 이상의 상호 독립된 집합으로 구분되는 선호도를 말한다. 비 이분적 선호도중 선호 단계가 무한히 많을 수 있는 경우를 연속적 선호도 (continuous preferences)라 한다.

자료검색질의는, 이상의 선호도 분류 기준에 따라, 이분적 검색질의와 비 이분적 검색질의로 분류된다. 이분적 검색질의는 이분적 선호도를 지니는 검색질의이며, 비 이분적 검색질의는 비이분적 선호도를 지니는 검색질의이다. 이분적 검색질의는 선호도의 이분성으로 인하여, 대상 실체에 대한 명확한 검색 기준으로 표시되며, 이러한 검색기준은 검색된 실체의 속성이 만족시켜야 할 명제들의 결합으로 표현된다. 이러한 점에서 이분적 검색질의를 정형적 검색질의라 칭한다. 전통적 자료검색은 기본적으로 정형적 (이분적) 검색질의를 가정한다.

비 이분적 검색질의는 선호도의 비 이분성으로 인하여 명확한 검색 기준의 기술 및 이에 대응하는 명제의 형성이 불가능하다. 선호도의 분류에서 언급된 바와 같이, 비 이분적 검색질의는 서로 다른

갯수의 선호 계급을 지니며, 이에 따라 서로 다른 특성을 지닐 수 있다. 본 연구는 비 이분적 검색질의 중 연속적 선호도를 지니는, 즉 선호 계급의 갯수가 무한히 커질 수 있는 검색질을 고려한다.

4. 자료검색의 분류

이상의 자료특성 분류와 검색요구 특성 분류에 근거하여 자료검색은 <표 1>에서 제시된 바와 같이 분류 요약된다.

전통적 자료검색은 완전 자료를 대상으로 이분적 검색질의에 의한 자료검색을 가정한다. 통상적으로 이러한 유형의 자료검색을 완전정보 검색 또는 정형적 자료검색이라 한다. 이에 반하여 데이터베이스가 부분자료이거나 누락속성을 포함할 경우를 불완전 정보 검색이라 하며, 검색질의가 비 이분적이며 선호계급의 수가 무한히 커질 수 있는 경우를 비 정형적 자료검색이라 한다.

본 연구는 비 정형적 자료검색을 위한 검색모형을 제시한다. 자료특성과 검색질의 특성에 기초하여 비 정형적 자료검색은 다음과 같이 정의된다.

<표 1> 자료검색 분류

자료 특성분류	완전자료	부분 자료	누락 속성
검색질의 특성분류			
이분적 검색질의	정형적 자료검색 (완전정보 검색)	불완전 정보 검색	
비이분적 검색질의	비정형적 자료검색	-	

정의 : 비 정형적 자료검색은 실체 e_1, e_2, \dots, e_n 을 선호 집단 E_1, E_2, \dots, E_k 로 분류하는 자료검색이다. 이때, 데이터베이스

는 물리적 실체에 대한 충분한 정보를 제공하며, n 과 k 는 무한히 클 수 있다.

Ⅲ. 비정형적 자료검색의 모형화

본 장은 비정형적 자료검색을 위한 근거모형을 제시한다. 비정형적 자료검색은 기존의 이분적 논리에 의하여 정형화될 수 없는 검색질의 및 검색기준을 어떻게 모형화 하느냐에 따라 각각 상이한 형태의 자료검색으로 모형화된다.

대표적인 비정형적 자료검색 모형으로 퍼지 자료검색모형 [Buckles and Petry, 1982; Umano, 1982; Baldwin and Zhou, 1984; Prades and Testemale, 1984], ARES [Ichikawa and Hirakawa, 1986], VAGUE [Motro, 1987] 등이 있다. 퍼지 자료검색모형은 Zadeh [1965]에 의하여 제시된 퍼지 집합이론에 근거하며, ARES와 VAGUE는 자료값 간의 유사성 (similarities)에 근거하고 있다.

본 연구는, 이들 기존의 모형과 독립적인 새로운 형태의 비정형적 자료검색모형으로서, 양기준 선호모형 (DAPM)을 제시한다.

1. 기초개념

본 논문에서 제시된 자료검색 모형은 근본적으로 선호이론에 기초한다. 즉, 자료검색은 정보이용자가 실체에 대하여 지니는 선호도의 측정으로 모형화된다. 그러나 전통적 선호이론에서 가정하는 수리적 선호함수는 자료검색시 이의 측정 및 적용이 용이하지 않다는 면에서 문제점을 지닌다 [Keeny, 1971]. 이들 문제점 해결을 위한 대안으로서, 본 연구는 자료실체의 속성값이 지니는 기하학적 특성에 기초한 새로운 형태의 선호모형을

제시한다.

선호모형 설계를 위한 기본 개념은 인간의 의사결정이 두 종류의 근거점 (anchors)에 기초하여 모형화될 수 있다는 것이다. 첫번째 근거점은 선택된 대안이 가능한 한 유사하기를 (가까이 위치하기를) 바라는 상상 가능한 최적의 이상적 대안이며, 두번째 근거점은 가능한 한 유사하지 않기를 (멀리 떨어져 위치하기를) 바라는 최악의 대안이다. 전자를 정 근거점 (positive anchor(s))이라 칭하며, 후자를 부 근거점 (negative anchor(s))이라 칭한다.⁴⁾

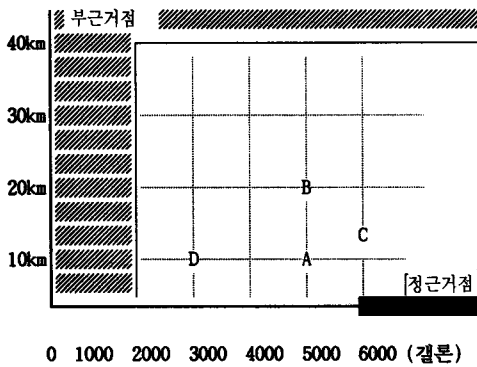
일 예로 앞에서 제시된 공장부지 선택을 고려하여 보자. 여기서 정 근거점은 6000갤론 이상의 공업용수 공급이 가능하며, 원자재 생산지에 접한 (거리가 0km 인) 부지이며, 부 근거점은 공업용수 공급이 2000갤론 미만이거나 원자재 생산지로부터 40km 이상 떨어져 전혀 고려될 수 없는 부지이다. 이들 근거점을 기준으로 하여 각 가용 부지에 대한 매입회사의 선호도는 해당부지의 속성이 정 근거점 또는 부 근거점과 유사한 (또는 상이한) 정도에 따라 결정된다.

〈표 2〉에 제시된 바와 같이 4개의 가용부지가 존재한다고 가정하자. 가용 부지 및 근거점의 기하학적 특성은 [그림 2]에 제시된 바와 같이 이차원 Euclidian 공간내에서 도형화된다. [그림 2]로부터 우리는 부지_A가 부지_B 또는 부지_D 보다 선호된다고 말할 수 있다 (부지_A가 부지_B 또는 부지_D 보다 정 근거점에 가까이, 부 근거점에 멀리 위치하는 것으로 “보이기” 때문에). 그러나 부지_A와 부지_C간의 선호관계는 직관적으로 결정될 수 없으며, 이들 각각이 정 근거점 및 부 근거점으로 부터 얼마나 가까이 위치하는가를 측정함으로써 결정된다.

4) 정 근거점은 Zeleny [1976]의 선택공리에서 제시된 근거점과 동일한 개념이다. 따라서 본 연구에서 제시된 선호모형은 Zeleny의 선택이론에 부 근거점을 추가하여 확대 보완한 것으로 해석될 수 있다.

〈표 2〉 가용부지의 속성값

	부지_A	부지_B	부지_C	부지_D
공업용수 공급량 (갤론)	5000	5000	6000	3000
생산지로 부터의 거리(km)	10	20	15	10



[그림 2] 가용부지 속성값의 기하학적 특성

이상에서 설명된 기본개념에 기초하여, 인간의 의사결정은 다음과 같은 선호공리에 의하여 개념화된다.

선호공리: 인간의 의사결정은 상상 가능한 최적의 이상적 대안인 정 근거점과, 최악의 대안인 부 근거점에 기초한다. 각 대안은 정 근거점과 가까울 (유사할) 수록, 부 근거점으로 부터 멀리 떨어질 (상이할) 수록 선호된다. 만약 전자가 후자를 보장하지 않는 경우에는 각 대안이 정 근거점 및 부 근거점에 상대적으로 가까이 위치하는 (유사한) 정도에 의하여 결정된다.

2. 모형화 : 양기준 선호모형 (DAPM: Double Anchors Preference Model)

제시된 선호공리에 근거하여 본 연구는 비 정형적 자료검색을 위한 의사결정모형으로서 양기준 선호모형(DAPM)을 제시한다. 선호공리의 적용과 관련하여, 본 연구는 아래의 조건이 만족됨을 가정한다.

- 각 개별 속성에 대한 정보이용자의 선호도는 최적치 (the most desirable value) 및 최악치 (the least desirable value)에 의하여 제한된다. 최적치는 기타 모든 속성값보다 선호되는 값이며, 최악치는 기타 모든 속성값에 비해 선호되지 않는 값이다.
- 정 근거점과 부 근거점은 각 속성별 최적치 및 최악치를 결합함으로써 정의된다. 이들 값은 절대치 (예: 6000 갤론) 또는 상대치 (예: 가용 부지의 공업용수 공급량중 최대치)로 표시된다.

비 정형적 의사결정을 위한 근거점의 설정 및 근거점으로 부터의 거리 (유사성) 측정은, 상기 선호공리 및 가정에 기초하여, 다음과 같이 정의된다.

2.1 근거점의 설정

가장 단순한 의사결정 문제로서 단일 속성 의사결정을 고려하여 보자. 이 경우 정 근거점은 대상속성의 값이 최적치인 실체 (또는 실체의 집합)로 정의되며, 부 근거점은 대상속성의 값이 최악치인 실체 (또는 실체의 집합)로 정의된다. 그러나 둘 이상의 속성을 동시에 고려할 경우, 근거점 설정은 이와 같이 간단하지 않으며, 속성간의 연관성에 의하여 다르게 유도된다.

상기 공장부지 선정 예에서 정 근거점은 공업용수 공급이 6000갤론 이상이며, 생산지로 부터 거리가 0 km인 부지들의 집합으로 정의되었으며, 부 근거점은 공업용수 공급이 2000갤론 미만이거나 생산지로 부터 거리가 40km 이상인 부지들의 집합으로 정의되었다. 이러한 근거점의 설정은 선정 기준으로 공업용수 공급량과 생산지로 부터의 거리가 서로 독립적이라는 점에 기인한다.

동일한 공장부지 선정예에서 선정기준이 공업용수 공급량과 (생산지로 부터의 거리가 아닌) 상수원으로 부터의 거리로 변경되었다고 가정하자. 상수원으로 부터의 거리가 고려된 이유는 공업용수가 부족한 경우 공업용수의 자체 확보를 위함이라고 가정할 경우, 정 근거점은 6000갤론 이상의 공업용수 공급이 가능하거나 또는 상수원에 접하여 (거리가 0km) 공업용수의 자체 확보가 가능한 부지들이며, 부 근거점은 공업용수 공급이 2000갤론 미만이며, 상수원으로 부터 40km 이상 떨어져 자체 공업용수의 확보가 전혀 불가능한 부지들이다.

이들 예가 제시하는 바와 같이, 의사결정의 근거점은 속성들간의 연관관계에 의하여 다르게 유도된다.

속성간 연관관계를 정의하는 접속어로서, 본 연구는 AND 와 OR를 고려한다. 개략적으로 이야기하여, AND는 상호독립적인 속성을 결합하며, OR은 상호보완적인 속성을 결합한다.

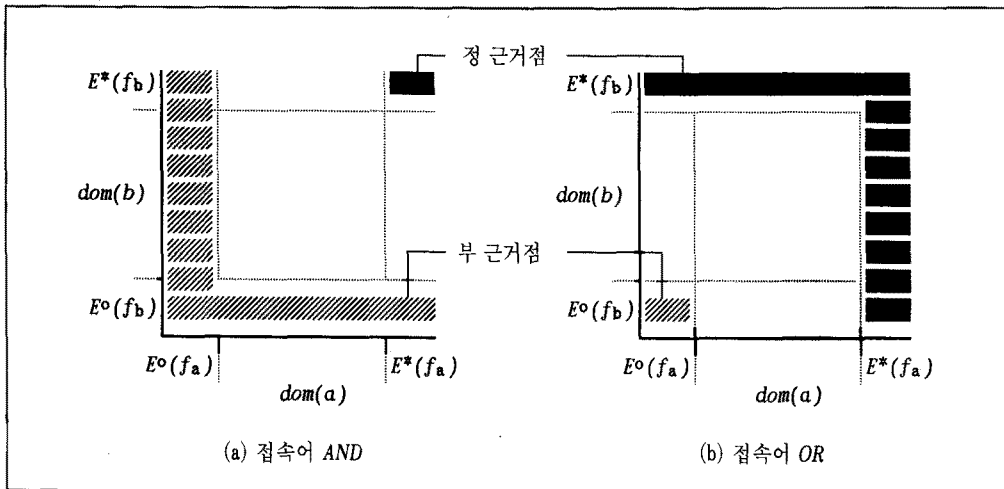
이들 접속어에 대응하는 근거점의 형성은 다음과 같이 정형화된다.

근거점 형성 법칙 : 속성 a 에 대한 선택기준을 fa , b 에 대한 선택기준을 fb 라 하자. 이들 fa 와 fb 에 대응하는 정 근거점을 각각 $E^*(fa)$, $E^*(fb)$, 부 근거점을 $E^o(fa)$, $E^o(fb)$ 라 하자. 속성 a 와 b 를 동시에 고려한 의사결정의 근거점은 접속어의 성격에 따라 다음과 같이 정의된다.

접속어 AND: $E^*(fa \text{ AND } fb) = E^*(fa) \times E^*(fb)$
 $E^o(fa \text{ AND } fb) = E^o(fa) \times \text{dom}(b) \cup \text{dom}(a) \times E^o(fb)$

접속어 OR: $E^*(fa \text{ OR } fb) = E^*(fa) \times \text{dom}(b) \cup \text{dom}(a) \times E^*(fb)$
 $E^o(fa \text{ OR } fb) = E^o(fa) \times E^o(fb)$

여기서 \times 는 Cartesian product이며, $\text{dom}()$ 는 해당 속성의 속성값 영역을 나타낸다.



[그림 3] 근거점 형성법칙

근거점 형성법칙에 의한 정 근거점 및 부 근거점은 [그림 3]에 예시된 바와 같다.

본 연구에서 제시된 근거점 형성 법칙의 주 특성은 반복적 (recursive)이라는 점이다. 즉, 상기 법칙은 AND와 OR에 의하여 결합된 임의의 검색 조건에 대하여 적용 가능하다.

따라서, 셋 이상의 속성을 포함하는 검색조건인 경우에도, 각 근거점은 근거점 형성법칙으로 부터 반복적으로 유도된다. 일 예로서 세 종류의 속성 a , b 및 c 를 포함하는 검색조건 $(fa \text{ AND } fb) \text{ OR } fc$ 를 고려하여

보자. 이 경우 각 근거점은 다음과 같이 유도 된다.

1 단계: 접속어 OR를 기준으로 (fa AND fb)를 첫 번째 검색조건으로 하고 fc 를 두번째 검색조건으로 취급하여 근거점 형성법칙을 적용한다.

$$\begin{aligned} & \text{(예) } E^*((fa \text{ AND } fb) \text{ OR } fc) \\ &= E^*(fa \text{ AND } fb) \times dom(c) \cup dom(a \text{ AND } b) \times E^*(fc) \\ & \quad E^0((fa \text{ AND } fb) \text{ OR } fc) \\ &= E^0(fa \text{ AND } fb) \times E^0(fc) \end{aligned}$$

2 단계: 접속어 AND를 기준으로 fa 를 첫번째 검색 조건으로 하고 fb 를 두번째 검색조건으로 취급하여 근거점 형성법칙을 적용함으로써 $E^*(fa \text{ AND } fb)$ 및 $E^0(fa \text{ AND } fb)$ 를 유도한다. 여기서 결합된 속성의 속성값 영역은 개별 속성값 영역의 Cartesian product로 정의된다. 즉,

$$\begin{aligned} dom(a \text{ AND } b) &= dom(a) \times dom(b) \\ &= dom(a \text{ OR } b) \text{ 이다.} \end{aligned}$$

$$\begin{aligned} & \text{(예) } E^*((fa \text{ AND } fb) \text{ OR } fc) \\ &= E^*(fa \text{ AND } fb) \times dom(c) \cup dom(a \text{ AND } b) \times E^*(fc) \\ &= E^*(fa) E^*(fb) \times dom(c) \cup dom(a) \times dom(b) \times E^*(fc) \\ & \quad E^0((fa \text{ AND } fb) \text{ OR } fc) \\ &= E^0(fa \text{ AND } fb) \times E^0(fc) \\ &= (E^0(fa) \times dom(b) \cup dom(a) \times E^0(fb)) \times E^0(c) \\ &= E^0(fa) \times dom(b) \times E^0(c) \cup dom(a) \times E^0(fb) \times E^0(c) \end{aligned}$$

이상의 예로부터 추론될 수 있는 바와 같이, 임의의 n 개의 속성을 포함하는 검색조건에 대하여서도, 각 근거점은 제시된 형성법칙을 반복적으로 적용함으로써 유도된다.

2.2 상대 거리 (Relative Distance)

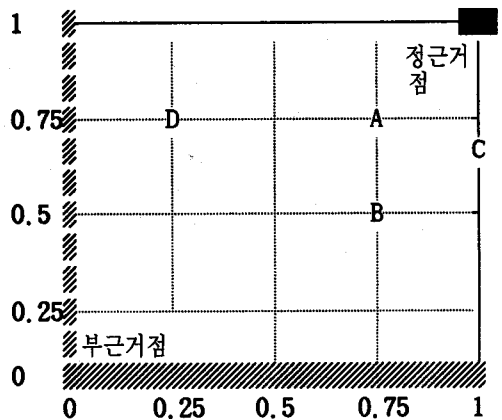
의사결정 선호공리에 의하면, 각 대안에 대한 선호도는 이들과 근거점과의 인접성 (유사성)으로 결정된다. n 개의 속성을 고려한 의사결정에 있어 정 근거점 및 부 근거점은 n 차원 Euclidian 공간상의 부분공간으로 표현되며, 각 대안은 n 차원 공간에서 하나의 점으로 표현된다. 따라서 이들간의 인접성은 n 차원 Euclidian 거리로 정형화된다.

그러나 n 차원 공간상의 거리 측정을 위하여서는 각 속성 측정단위간의 상이성 (예: 공업용수 공급량은 갤런, 거리는 km)이 먼저 해결되어야 한다. 이를 위하여, 본 연구는 각 속성값을 0부터 1까지의 실수값으로 변환한 정규 속성값을 제시한다.

정규 속성값(v^*)은 해당 속성의 최적치(c^*)가 1의 값을 지니며, 최악치(c^0)가 0의 값을 지니도록 변환한 값이다. 즉, 정규 속성값 v^* 는 속성값 v 로 부터 다음과 같이 유도된다.

$$v^* = \frac{(v - c^0)}{(c^* - c^0)}$$

정규 속성값에 의하면 각 속성의 속성값 영역은 $[0, 1]$ 로 정의되며, n 차원 Euclidian 공간은 각 축이 $[0, 1]$ 의 값을 지니는 정규 n 차원 공간으로 변환된다 [그림 4] 참조).



[그림 4] 가용부지 정규 속성값의 기하학적 특성

정규 n 차원 Euclidian 공간에서 각 대안에 대한 의사결정자의 선호도는 이들과 근거점과의 거리를 상호 비교함으로써 유도된다. 일 예로 대안 ei 및 ej 를 고려하여 보자. 이들 대안과 정 근거점 E^* 및 부근거점 E^0 과의 거리를 각각 $d(ei, E^*)$, $d(ei, E^0)$ 및 $d(ej, E^*)$, $d(ej, E^0)$ 라 할 경우, 대안 ei , ej 간의 선호관계는 다음과 같이 유도된다:

$d(ei, E^*) \leq d(ej, E^*)$ 와 $d(ei, E^0) \geq d(ej, E^0)$ 의 두 조건이 동시에 만족되면, ei 는 ej 에 대하여 선호된다 (역도 성립).

그러나, $d(ei, E^*) \leq d(ej, E^*)$ 와

$d(ei, E^0) \geq d(ej, E^0)$ 의 두 조건 중 하나만 만족될 경우에는

(즉, $d(ei, E^*) \leq d(ej, E^*)$ 이고
 $d(ei, E^0) \leq d(ej, E^0)$ 이거나,
 $d(ei, E^*) \geq d(ej, E^*)$ 이고
 $d(ei, E^0) \geq d(ej, E^0)$ 인 경우)

두 대안 ei 와 ej 간의 선호관계는 정 근거점 및 부근거점으로 부터의 거리를 나타내는 $d(ei, E^*)$, $d(ei, E^0)$ 및 $d(ej, E^*)$, $d(ej, E^0)$ 로부터 직접 유도될 수 없다.

이러한 선호관계의 비 결정성을 해결하기 위하여, 본 연구는 상대거리 (relative distance) 개념을 도입한다. 상대거리란 각 대안이 부근거점을 기준으로 하여 정 근거점으로 부터 어느 정도 떨어져 있는가를 측정하는 단위로서 다음과 같이 유도된다.

$$rd(e; E^*, E^0) = \frac{d(e, E^*)}{d(e, E^*) + d(e, E^0)}$$

여기서 $rd(e; E^*, E^0)$ 는 대안 e 의 근거점 E^* 와 E^0 로 부터의 상대거리를 의미한다.

의사결정 선호모형은 상대거리를 이용하여 다음과 같이 실 적용 가능 공리로 재 정의된다.

임의의 두 대안 ei 와 ej 에 대하여, 만약

$$rd(ei; E^*, E^0) \leq rd(ej; E^*, E^0) \text{이면,}$$

ei 는 ej 에 대하여 선호된다.

일반적 선호모형에 있어 선호값은 클수록 높은 선호도를 나타낸다. 이러한 일반적 표현기준을 따르기 위하여, 양기준 선호모형에 의한 선호값은 다음과 같이 정의된다:

$$g(e) = 1 - rd(e; E^*, E^0) \\ = \frac{d(e, E^0)}{d(e, E^*) + d(e, E^0)}$$

여기서 $g(e)$ 는 실체 e 의 양기준 선호모형에 의한 선호값을 나타낸다.

2.3 양기준 선호모형의 적용 예

본 절에서는 이상에서 제시된 양 기준 선호모형의 실 적용을 예시한다. 실 적용 예로서 <표 2>에 제시된 가용부지의 검색 예를 고려하자. 이 경우, 공장 부지에 대한 선호도는 <표 3>에 제시된 바와 도출된다.

부지간의 선호도는 양기준 선호모형의 기초개념에서 개략적으로 예측한 바와 일치한다. 즉, 부지_A는 부지_B 또는 부지_D 보다 정 근거점에 가까이, 부근거점에 멀리 위치하기 때문에 부지_A는 부지_B 또는 부지_D 보다 높은 선호값을 지니도록 계산되었다.

또한 직관적으로 결정될 수 없는 부지_A와 부지_C 또는 부지_B와 부지_D에 대하여서는 부지_A는 부지_C보다 선호되는 것으로 부지_B는 부지_D보다 선호되는 것으로 계산되었다. 이러한 결과는 부지_A와 부지_B 각각이 부지_C와 부지_D 보다 상대적으로 정 근거점에 가까이 위치하기 때문이다.

〈표 3〉 양 기준 선호모형 적용 예: 공장부지 선호도의 결정

	공업용수 공급량 (갤론)	거리 (km)	정규속성값 (공업용수공급량, 거리)	$d(e, E^*)$	$d(e, E^0)$	$d(e, E^*)$ + $d(e, E^0)$	$g(e)$	순위
부지_A	5000	10	(0.75, 0.75)	0.35	0.75	1.10	0.68	1
부지_B	5000	20	(0.75, 0.50)	0.56	0.50	1.06	0.47	3
부지_C	6000	15	(1.00, 0.625)	0.375	0.625	1.00	0.625	2
부지_D	3000	10	(0.25, 0.75)	0.79	0.25	1.04	0.24	4

(주) 본 예에서 정 기준점 및 부 기준점은 다음과 같이 표현된다.

$$E^* = (1, 1), E^0 = \{ (x, 0) (0, y) \mid 0 \leq x \leq 1, 0 \leq y \leq 1 \}$$

IV. 논리적 특성

본 절은 양 기준 선호모형이 자료검색 모형으로서 지니는 타당성을 분석 평가한다. 이를 위하여 본 연구는 양기준 선호모형의 논리적 특성을 분석 제시한다.

자료검색 모형이 지녀야 할 특성의 기본 근거는 부울 논리(boolean logic)이다. 완전 대응 (exact matching) 자료검색 모형의 근거 이론인 부울 논리는 논리적 단순성, 명료성과 더불어 검색 결과의 완전성 및 일관성을 충족시키는 것으로 증명되고 있다 [Reiter, 1984]. 그러나 비정형적 자료검색을 위하여서는 일반적으로 공인되는 논리적 근거이론이 존재하지 않는다.

그러나, 자료검색의 분류 정의에서 제시된 바와 같이, 정형적 자료검색은 비정형적 자료검색의 특수 경우이다. 따라서 바람직한 비 정형적 자료검색 모형은 이의 특수 경우인 정형적 자료검색 또한 지원할 수 있어야 한다. 이러한 비 정형적 자료검색의 특성으로 인하여 정형적 자료검색의 근거이론인 부울 논리는 비 정형적 자료검색의 특성 분석을 위한 최소 기준으로 이용될 수 있다.

자료검색과 관련하여 부울 논리의 기본법칙은 〈표 4〉에 제시된 바와 같이 요약된다. 이들 기본

법칙에 대하여, 양기준 선호모형은 중간점 제외법칙 (Law of Excluded Middle)을 제외한 제 법칙을 충족시킴이 증명되었다 [Lee,1990]. 이러한 특성을 논리적 관점에서 재 정리하면, 양 기준 선호모형에 의한 자료검색은 드모르간 논리(DeMorgan Logic)를 따름을 의미한다.

양기준 선호모형이 드모르간 논리를 따름은 이에 기초한 자료검색이 다음과 같은 논리적 특성을 지님을 의미한다.

첫째, 모형에 의한 비정형적 검색질의의 해석 결과가 완전성 및 일관성을 충족시킨다.

둘째, 정형적 자료검색질의의 해석과 관련하여 완전대응 자료검색 결과와의 비교 근거를 제공한다.

구체적으로 부울 논리는 연산수(operand)가 0 또는 1의 값을 가지는 디모간 논리의 특수 경우이다. 따라서, 양기준 선호모형이 드모르간 논리를 따름은, 개별 속성에 대한 정형적 검색질의의 해석이 0 또는 1의 값을 지닐 경우, 이들 속성을 결합한 모든 (정형적) 검색질의의 해석이 부울 논리를 따름을 의미한다.

개별 속성에 대한 정형적 검색질의의 해석 값은 다음에 제시된 바와 같이 0또는 1의 값을 지닌다. (증명: 정형적 검색질의의 해석과 관련하여 양기준 선호모형의 각 근거점은 상호 배제적이다. 즉 속성

a의 검색기준 fa에 대한 정 근거점 및 부 근거점을 각각 $E^*(fa)$, $E^0(fa)$ 라 할 경우,

$$E^*(fa) \cup E^0(fa) = \text{dom}(a) \text{ 이며}$$

$$E^*(fa) \cap E^0(fa) = \emptyset \text{이다.}$$

이로부터 $d(e, E^*)$ 와 $d(e, E^0)$ 중 하나는 0이며 따라서 $d(e; E^*, E^0)$ 는 0 또는 1의 값을 지닌다.

이상의 결과를 종합할때 양기준 선호모형은 첫째, 자료 검색의 완전성 및 일관성을 충족시키며, 둘째, 전통적 완전대응 자료검색 모형을 포함한다. 즉, 정형적 자료검색 질의에 대하여 양기준 선호모형은 완전대응 자료검색 모형과 동일한 결과를 산출한다. 이러한 분석 결과는 제안된 양기준 선호모형이 자료검색을 위한 의사결정 모형으로서 바람직한 제 특성을 보유하고 있음을 증명한다.

〈표 4〉 부울 논리의 기본법칙

1. 항진 법칙 (Principle of Tautology)

$$(f \wedge f)(e) = f(e)$$

$$(f \vee f)(e) = f(e)$$

2. 교환 법칙 (Principle of Commutation)

$$(f \wedge g)(e) = (g \wedge f)(e)$$

$$(f \vee g)(e) = (g \vee f)(e)$$

3. 결합 법칙 (Principle of Association)

$$(f \wedge (g \wedge h))(e) = ((f \wedge g) \wedge h)(e)$$

$$(f \vee (g \vee h))(e) = ((f \vee g) \vee h)(e)$$

4. 배분 법칙 (Principle of Distribution)

$$(f \wedge (g \vee h))(e) = ((f \wedge g) \vee (f \wedge h))(e)$$

$$(f \vee (g \wedge h))(e) = ((f \vee g) \wedge (f \vee h))(e)$$

5. 흡수 법칙 (Principle of Absorption)

$$(f \wedge (f \vee g))(e) = f(e)$$

$$(f \vee (f \wedge g))(e) = f(e)$$

6. 중간점제의 법칙 (Principle of Excluded Middle)⁵⁾

$$(f \wedge \neg f)(e) = 0$$

$$(f \vee \neg f)(e) = 1$$

V. 결론

본 연구는 종래의 이분적 자료검색 모형이 지니는 단점을 보완 대체하기 위한 대안으로서 양기준 선호모형을 제시하였다. 이를 위하여 본 논문은 첫째, 자료검색의 일반적 특성을 정의 분류하고, 둘째, 이들 일반적 자료검색 기준을 기술, 정의하기 위한 모형으로서 양기준 선호모형을 제시하였으며, 끝으로 제시된 모형의 논리적 특성을 분석하였다.

본 논문에서 제시된 양기준 선호모형은 다음의 두가지 측면에서 의미를 지닌다. 첫째, 새로운 형태의 선호모형으로서의 의미와, 둘째, 비 정형적 자료검색을 위한 의사결정 모형으로서의 의미이다. 후자와 관련하여 양기준 선호모형은 자료 검색의 완전성 및 일관성을 충족시키며, 또한 전통적 검색 질의에 대하여서는 완전대응 자료검색 모형과 동일한 결과를 산출한다. 이들 분석 결과는 제안된 선호모형이 자료검색을 위한 의사결정 모형으로서 바람직한 제 특성을 보유하고 있음을 증명한다.

양기준 선호 모형이 지니는 새로운 형태의 선호 모형으로서의 의미는 모형의 범용적 적용가능성의 분석이라는 면에는 중요성을 지닌다. 그러나 이에 대한 분석은 본 연구의 중점 사항으로 취급되지 않았으며, 따라서 추가적인 연구를 필요로 한다.

비정형적 자료검색을 위한 본 연구 결과의 실 적용을 위하여서는 이에 기초한 자료검색 시스템의 설계 및 구현을 위한 노력이 요구된다.

5) 연결자 - (NOT)을 포함한 검색질의의 해석과 관련하여, 정 근거점 및 부 근거점은 -을 포함하지 않는 검색질의의 근거점을 치환함으로써 설정된다.

즉, $E^*(-f) = E^0(\neg f)$, $E^0(\neg f) = E^*(f)$

로 정의된다 ([LEE, 1990] 참조).

이를 위하여 [LEE, 1990]은 기존의 데이터베이스 검색언어 (예: SQL)에 기초한 검색 시스템을 제시하였다. [LEE, 1990]에 의하여 제시된 검색 시스템은 비 정형적 검색질의의 기술 및 기준점의 설정, 상대거리 측정 모듈들을 기존 검색언어의 상위 시

스템으로 구현하고 있다. 그러나 자료검색 시스템의 기술적 효율성 (예: 검색 시간)을 향상시키기 위하여서는 양기준 선호모형을 위한 독자적인 검색언어의 설계 및 구현이 요구되며, 이를 위한 추가적인 연구가 필요하다.

참고 문헌

- ANSI/X3/SPARC. "Study Group on Database Management Systems: Interim Report", *FDT (Bulletin of ACM SIGFIDET)*, 7(2), February 1975. Baldwin, J. F. and, S. Q. Zhou. "A Fuzzy Relational Inference Language", *Fuzzy Sets and Systems*, 14, 1984, pp.155-174.
- Buckles, B. P. and, F. E. Petry. *Fuzzy Databases and Their Applications*, in *Fuzzy Information and Decision Processes*, M. M. Gupta and E. Sanches (eds.), North Holland Publishing Company, 1982, pp.361-371
- Ichikawa, T. and, M. Hirakawa. "ARES: A Relational database with the Capability of Performing Flexible Interpretation of Queries", *IEEE Transactions on Software Engineering*, SE 12(5), May 1986, pp.624-634.
- Keeny, R. L. "Utility Independence and Preference for Multiattributed Consequences", *Operations Research*, 19, July 1971, pp.875-893.
- Lee, C. Y., *A Multi-criteria Data Retrieval Model: An Application of Multi-attribute Preference Model to data Retrieval*, unpublished Ph. D. Thesis, University of Michigan, Ann Arbor, MI, 1990.
- Lipski, Jr., W. "On Semantic Issues Connected with Incomplete Information Databases", *ACM Transactions on Database Systems*, 4(3), 1979, pp.262-296.
- Lyytinen, K. "Two views of Information Modeling", *Information and Management*, 12, 1987, pp. 9-19.
- Motro, A. "VAGUE: A User Interface to Relational Databases that Permit Vague Queries", *ACM Transactions on Office Information Systems*, 6(3), July 1988, pp.187-224.
- Prade, E. and, C. Testemale. "Generalizing Database Relational Algebra for the Testment of Incomplete or Uncertain Information and Vague Queries", *Information Science*, 34, 1984, pp. 115-143.
- Reiter, R. *Towards a Logical Reconstruction of Relational database Theory*, in *On Conceptual Modeling*, M Brodie and, J. Mylopoulos and, J. W. Schmidt (eds.), Springer-Verlag, New York, 1984.

Umano, M. *Freedom-O: A Fuzzy Database Systems*, in *Fuzzy Information and Decision Processes*, M. M. Gupta and E. Sanches (eds.), North Holland Publishing Company, 1982, pp.339-347

Information in Database Systems”, *ACM Transactions on Database Systems*, 7(3), pp.470-488.

Zadeh, L. A. “Fuzzy Sets”, *Information and Control*, January 1965, pp.338-353.

Wong, E. “A Statistical Approach to Incomplete

◇ 저자소개 ◇



저자 이춘열은 한국통신 소프트웨어연구소에 재직 중이다. 그는 서울대학교 산업공학과를 졸업하고 동 대학원 경영학과에서 경영학 석사학위를 취득하였다. 또한 국방관리연구소 전산과학연구위원회 (현 국방정보체계연구소)에서 국방정보체계의 설계 및 분석 평가에 참여하였다. 그 후 미시간 대학교 (앤아버 미시간)에서 경영정보학을 전공하여 경영학 박사학위를 취득하였다. 그의 주요 관심분야는 데이터베이스 응용, 자료검색, 정보체계의 개발관리 및 정보공학등이다.