

# 질의 확장에 의한 데이터베이스 검색

박찬영<sup>u</sup>                      김정호                      정 흥  
계명대학교 컴퓨터공학과  
cypark@keobuksun.keimyung.ac.kr  
phjhkim@hanmail.net  
jhong@kmu.ac.kr

## Retrieval of Databases Using Query Extension

Chan-Young Park<sup>U</sup>                      Jung-Ho Kim                      Hong Chung  
Dept. of Computer Engineering, Keimyung University

### 요 약

데이터베이스에 대하여 아무런 지식이 없는 일반인도 데이터베이스를 쉽게 검색할 수 있도록 언어변수를 사용한 질의 및 질의 확장에 의한 효율적인 데이터베이스 검색 시스템을 설계한다. 언어 변수의 퍼지화 및 질의 확장을 위해 퍼지 소속함수, 개념 계층, 유사 관계 등을 적용하며, 한의원 데이터베이스를 사례로 하여 프로토타입을 구현하고 실험 및 평가를 한다.

### 1. 서론

현재 정보화 사회를 구성하는데 있어서 핵심적인 역할을 하는 분야 중 하나가 데이터베이스인데, 점차 다양화되고 활용도가 높아짐에 따라 일반인들도 많이 이용하고 있다. 그러나 데이터베이스에 대한 지식이 없는 일반인들이 필요한 정보를 유효하게 검색하기에는 아직 질의에 대한 어려움이 많으며, 단순한 질의만으로는 원하는 정보를 찾기가 쉽지 않다. 일반인들이 쉽게 정보검색을 하도록 하려면 언어변수를 사용하여 질의를 할 수 있도록 해야 하며, 또한 원하는 검색이 되지 않을 때는 질의 내용을 완화 내지는 확장하여 근사적인 검색이 가능하도록 할 필요가 있다[1]. 따라서 본 연구에서는 퍼지 소속함수 [2,6]와 개념 계층[3]을 사용하여 언어변수로 질의를 할 수 있도록 하며, 퍼지 소속함수, 개념 계층, 유사 관계[8]를 사용하여 질의 확장을 할 수 있도록 함으로써 근사 검색이 가능한 데이터베이스 검색 시스템을 설계하고자 한다. 그리고 한의원 데이터베이스를 사례로 하여 프로토타입 시스템을 구현하고 실험 및 평가를 한다.

### 2. 관련 이론

본 장에서는 언어 변수의 퍼지화를 위한 퍼지 소속함수, 질의 내용의 완화에 사용될 개념 계층과 유사 관계에 대하여 간단히 고찰해 보고자 한다.

#### 2.1 퍼지 소속함수

퍼지 집합은 "큰 키", "중년" 등과 같이 경계가 확실하지 않은 애매 모호한 성질을 표현하는 집합으로서, 퍼지 집합을 이루는 실수의 단조증가의 정의역을 나타내는 횡축, 소속정도 [0,1]을 나타 내는 종축으로 표현하는 좌표 상에서 정의역과 소속함수를 커브로 연결하여 퍼지 집합을 이룬다[2].

소속함수는 일반적으로 선(Linear)형, S(Sigmoid)형, 벨(Bell)형으로 구분된다.

• 선형함수: 퍼지 집합의 모양이 선형으로서 소속함수가 0이 되는 값을  $\alpha$ , 1이 되는 값을  $\beta$ 라 할 때, 다음과 같이 정의된다.

$$L(x; \alpha, \beta) = \begin{cases} 0 & : x \leq \alpha \\ (x - \alpha) / (\beta - \alpha) & : \alpha < x < \beta \\ 1 & : x \geq \beta \end{cases}$$

감소형은  $1 - L(x; \alpha, \beta)$ 이다. 선형 함수는 "크다", "낮다" 등과

같이 비례적인 의미를 가진 언어변수의 표현에 주로 사용된다. S형 함수: 퍼지 집합의 모양이 비선형으로서, 소속함수가 0이 되는 값을  $\alpha$ , 1이 되는 값을  $\gamma$ , 0.5가 되는 변곡점을  $\beta$ 라 할 때, 다음과 같이 정의된다.

$$S(x; \alpha, \beta, \gamma) = \begin{cases} 0 & : x \leq \alpha \\ 2((x - \alpha) / (\gamma - \alpha))^2 & : \alpha < x \leq \beta \\ 1 - 2((x - \gamma) / (\gamma - \alpha))^2 & : \beta < x < \gamma \\ 1 & : x \geq \gamma \end{cases}$$

감소형은  $1 - S(x; \alpha, \beta, \gamma)$ 이다. S형 함수는 "빠르다", "젊다" 등과 같이 가속적인 의미를 가진 언어 변수의 표현에 주로 사용된다.

• 뿔형: 퍼지 집합의 모양이 뿔 모양으로서, PI형, Beta형, Gauss형 등이 있는데, 대표적으로 가장 많이 사용되는 것이 PI형이다. 이는 증가 S형과 감소 S형을 합친 형태로 다음과 같이 정의된다.

$$\Pi(x; \beta, \gamma) = \begin{cases} S(x; \gamma - \beta, \gamma - \beta/2, \gamma) & : x \leq \gamma \\ 1 - S(x; \gamma, \gamma + \beta/2, \gamma + \beta) & : x > \gamma \end{cases}$$

$\gamma$ : 정의역의 중앙값

$\beta$ : 중앙값으로부터 소속도가 0이 되는 곳까지의 폭

PI형 함수는 "중년", "5정도" 등과 같이 퍼지량을 표현하는데 주로 사용된다.

#### 2.2 개념계층

개념계층은 데이터베이스 속성에 대한 일반화 관계의 집합으로서 개념들을 계층적인 형태 또는 부분정렬로 구성하여 지식을 간결하게 또는 상위 단계의 용어들로 표현 하기 위해 사용된다. 일반화 관계는  $1 \leq i \leq k$ 에 대한 각각의  $A_i$ 를 일반화한  $B$ 를  $\{A_1, A_2, \dots, A_k\} \subset B$ 로 표현할 수 있다[3].

개념계층을 생성하는 방법에 따라 스키마 계층, 셋-묶음 계층, 연산 기반 계층, 규칙 기반 계층[4,5] 등이 있는데 본 시스템에서 사용할 셋-묶음 계층에 대해서만 간단히 기술한다.

이는 데이터베이스의 속성 값 집합에 대한 관계성을 묶어 하나의 집합으로 정의함으로써 주어진 응용범주에서의 의미론적인 관계성을 반영하는 구조를 말한다. 예를 들어, 대학교의 학생 계층을 다음과 같은 개념계층으로 나타낼 수 있다.

$$\{\text{freshman, sophomore, junior, senior}\} \subset \text{undergraduate} \subset \text{graduate} \subset \text{M.S, M.A, Ph.D}$$

2.3 유사 관계

유사 관계는 불확실한 데이터베이스와 질의어를 처리하기 위한 퍼지 데이터베이스[8]에서 질의 확장을 위해 사용되는 한 방법이다. 예를 들어 표-1과 같은 "직업"에 대한 유사관계가 있다고 하자.

표-1 유사관계

직업	교수	교사	의사	변호사	목사
교수	1	0.8	0.2	0.3	0.5
교사	0.8	1	0.3	0.5	0.8
의사	0.2	0.3	1	0.7	0.3
변호사	0.3	0.5	0.7	1	0.6
목사	0.5	0.8	0.3	0.6	1

예를 들어 직업이 교수인 사람을 찾고자 하는데 없다고 하자. 이때 질의를 완화하여 유사정도가 0.7이상인 사람이라고 하면 직업이 교사인 사람도 검색이 될 것이다. 물론 이때의 검색 결과는 근사 검색이다.

유사 관계 외에 퍼지 데이터베이스 모델인 VAGUE 모델[7]에서 사용하는 거리표가 있으나 유사정도를 역으로 나타냈을 뿐 동일한 내용이다.

3. 시스템 설계

3.1 시스템의 구조

본 연구에서는 한의원 데이터베이스를 사례로 하여 시스템을 설계하고자 한다. 시스템의 구조는 그림-1과 같다.

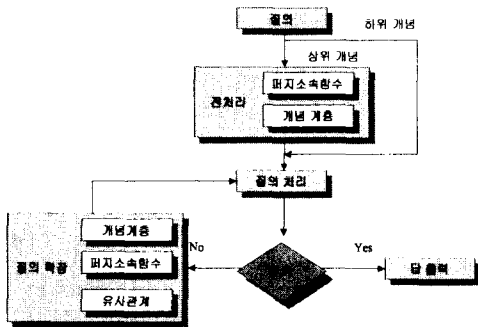


그림-1 시스템 구조도

이중 중요한 기능에 대하여 간략히 기술한다.

- 전처리: 질의 화면에 입력된 언어변수를 SQL문으로 변환한다. 이때 하위 개념의 질의일 경우는 바로 데이터베이스에 접근하나 상위개념의 질의일 경우는 퍼지 소속함수나 개념 계층을 거쳐 변환된다.
- 질의 확장: 개념계층, 유사관계 혹은 퍼지 소속도를 이용하여 질의를 완화 내지 확장한다. 이는 만족한 답이 나올 때까지 질의 확장을 반복한다.

3.2 전처리

본 한의원 데이터베이스 검색 시스템에서 질의 내용은 회원번호, 이름 이외에 언어변수로 표현되는 나이, 환자의 위치(주소), 초진날짜, 직업, 병명 등이 있다. 이중 SQL문을 위한 전처리 항목은 나이, 위치, 초진날짜이다.

- 나이는 "어린", "젊은", "중년", "노년"을 입력하는데 이때 사용자가 정의하는 소속도도 같이 입력한다. 이를 퍼지 집합으로 변환시킬 때, 그림-2와 같이 "어린"은 감소 S형 함수, "젊은"과 "중년"은 벨형 함수, "노년"은 증가 S형 함수를 사용한다.

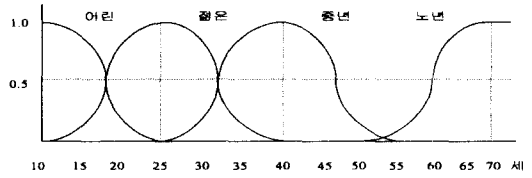


그림-2 나이 퍼지집합

- 위치는 다음 예와 같이 병원에 가까운 순서로 "근거리", "중거리", "원거리"로 표시하며, 각 거리별로 개념계층을 사용하여 동명으로 변환한다. 필요에 따라 동명을 입력할 수 있도록 한다.

- 근거리 ≡ {대명동, 봉덕동, 송현동, 이천동, 남산동, 중동}
- 중거리 ≡ {대당동, 본리동, 월성동, 상인동, 감삼동, 진천동}
- 원거리 ≡ {화원, 이곡동, 다사, 칠곡, 무태동, 방천동, 기타}
- 초진날짜는 "근래", "보통", "옛날"로 입력하는데 이때 사용자가 정의하는 소속도도 같이 입력한다. 이를 퍼지 집합으로 변환시킬 때 그림-3과 같이 "근래"는 감소 선형함수, "보통"은 삼각형 함수, "옛날"은 증가형 선형함수를 사용한다.

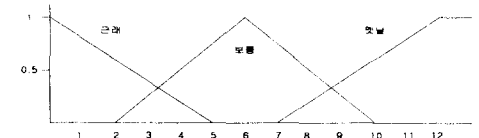


그림-3 초진날짜 퍼지집합

3.3 질의 확장

만족한 검색이 되지 않을 때는 질의 확장을 하는데, 입력 항목중 나이, 초진 날짜, 병명, 직업, 위치에 관하여 질의 확장을 한다. 질의 항목에 따라 확장 순서를 표시하면 순서에 따라 우선적으로 질의를 확장한다.

- 나이와 초진 날짜는 소속도 완화에 의하여 질의를 확장하는데, 각각 나이 퍼지집합, 초진 날짜 퍼지집합을 사용하여 확장된 SQL문을 재구성한다.
- 병명과 직업은 유사관계를 이용하여 가장 유사한 병명과 직업으로 질의를 변환한다.
- 위치는 개념계층을 이용하여 가장 유사한 위치로 질의를 변환한다.

4. 시스템 구현 및 실험

본 시스템은 한의원 데이터베이스를 사례로 하여 VB 6.0으로 구현하며 데이터베이스는 Access를 사용한다.

4.1 전처리

예를 들어 "나이가 중년(소속도 0.5), 위치가 근거리, 초진날짜가 근래(소속도 0.7), 직업이 교사이고, 한요통 치료를 받은 사람"을 찾고 싶을 때, 질의 입력화면은 그림-4와 같다.

- 이 질의를 SQL문으로 변환하는 알고리즘은 다음과 같다.
- 나이는 "어린", "젊은", "중년", "노년"에 따라 각각의 퍼지 집합을 소속도로 alpha-cut하고, x좌표 값의 범위로 치환한 후 생년월일로 변환;
- 위치는 "근거리", "중거리", "원거리"에 따라 위치 개념계층에서 동명으로 치환함;
- 초진날짜는 "근래", "보통", "옛날"에 따라 각각의 퍼지 집합을 소속도로 alpha-cut하고, x좌표 값의 범위로 치환한 후 연월일로 변환;

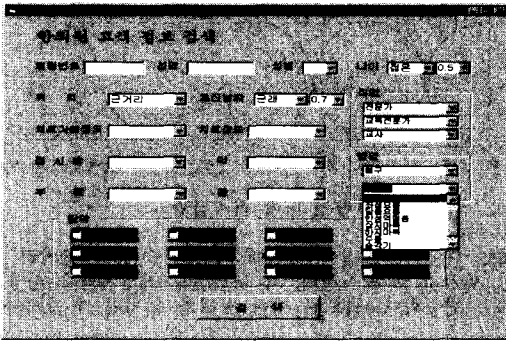


그림-4 질의 입력 화면

따라서 전처리된 SQL문은 다음과 같이 된다.

```
Select * From Client_table
Where (dob >= 19520615 And dob <= 19670715)
And (pos = "대명동" Or pos = "봉덕동" Or .....)
And (dat >= 20000106) And (직업 = "교사")
And (병명 = "한요통");
```

4.2 질의 확장

만족한 검색이 되지 않아 예를 들어 그림-5와 같은 질의 확장을 했다고 하자.

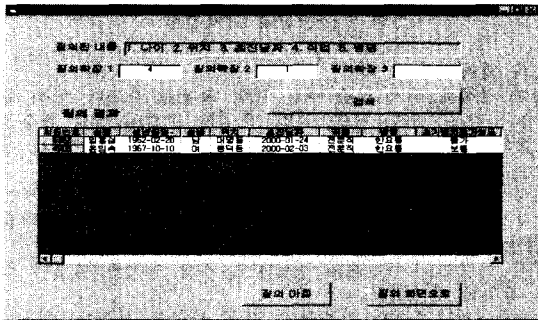


그림-5 질의 확장 화면

이 질의확장을 SQL문으로 변환하는 알고리즘은 다음과 같다.

- 나이에 대해서는 소속도에서 0.1을 감하여 나이 퍼지집합을 alpha-cut하고, x좌표 값의 범위로 치환한 후 생년월일로 변환;
- 초진날짜에 대해서는 소속도에서 0.1을 감하여 초진날짜 퍼지집합을 alpha-cut하고, x좌표 값으로 치환한 후 연월일로 변환;
- 병명에 대해서는 병명 유사관계에서 유사도가 가장 큰 병명으로 치환;
- 직업에 대해서는 직업 유사관계에서 유사도가 가장 큰 직업으로 치환;
- 위치에 대해서는 위치 개념계층에서 동일 그룹에 속하는 동명으로 치환;

따라서 변환된 SQL문은 다음과 같이 된다.

```
Select * From Client_table
Where (dob >= 19520115 And dob <= 19671215)
And (pos = "대명동" Or pos = "봉덕동" Or .....)
And (dat >= 20000106) And (직업 = "전문직")
And (병명 = "한요통");
```

한의원 데이터베이스로부터 질의 확장에 의한 검색 결과는 그림-5와 같다.

5. 결론 및 평가

본 연구에서 실험적으로 구현한 질의확장에 의한 데이터베이스 검색 시스템은 질의 과정에서 언어변수를 사용한 개념적 질의가 가능하며, 만족한 검색을 하지 못했을 경우는 질의를 완화 내지는 확장하여 가장 가까운 근사 검색을 할 수 있다.

일반 데이터베이스의 질의와 비교할 때 다음과 같은 장점이 있다.

- 질의를 언어변수를 사용하여 할 수 있으므로 데이터베이스에 지식이 없는 일반 사용자도 쉽게 사용할 수 있다.
- 모호한 질의에 대한 검색이 가능하다.
- 원하는 답이 없을 경우 확장된 질의에 의해 가장 유사한 답을 찾을 수 있다.
- 질의가 확장됨으로써 검색에 실패가 적다.
- 반면 다음과 같은 문제점도 있다.
- 구현시 퍼지집합, 개념계층, 유사관계의 작성이나, 질의 입력시 퍼지 집합의 소속도 결정에 주관적인 요소가 있다.
- 전처리 및 질의확장 단계를 거쳐 데이터베이스를 검색하므로 검색속도가 느려질 수 있다.

전자에 있어서 모호한 질의를 처리한다는 의미에서 볼 때 피할 수 없는 문제점이지만 퍼지집합, 개념계층, 유사관계를 작성할 때는 영역 전문가의 도움을 받아서 구현하고, 사용시 소속도 결정에 있어서는 사용자가 퍼지 개념에 대해 이해를 하게 되면 별 문제가 되지 않을 것이다. 후자에 있어서도 검색 실패가 거의 없으므로 시스템의 전반적인 운용면에서 보면 그리 큰 문제가 되지는 않을 것이다.

본 프로토타입 시스템은 일부 항목에 대해서만 질의 확장을 했지만 기타 항목에 대해서도 언어변수를 분석하여 확장 적용이 바람직하다.

6. 참고문헌

- [1] W. W. Chu, H. Yang, K. Chiang, M. Minock, G. Chow, and C. Larson, "CoBase: A scalable and extensible cooperative information system". *Journal of Intelligent Information Systems*, Vol. 6, No. 2/3, pp. 223-259, 1996
- [2] Earl Cox, *Fuzzy Systems Handbook*, Prentice-Hall, 1994
- [3] D. Fudger and H. J. Hamilton, "A Heuristic for Evaluating Databases for knowledge Discovery with DBLEARN", *Rough Sets, Fuzzy Sets and Knowledge Discovery*, RSKD'93, pp. 44-51, 1993
- [4] J. Han, Y. Cai, and N. Cerecone, "Knowledge Discovery in Databases: An Attribute-Oriented Approach", *Proceedings of the 18th International Conference on VLDB*, Vancouver, URL: ftp://ftp.fas.sfu.ca/pub/cs/han/kdd/, 1992
- [5] J. Han and Y. Fu "Dynamic Generation and Refinement of Concept Hierarchies for Knowledge Discovery in Databases", URL:ftp://ftp.fas.sfu.ca/pub/cs/han/kdd/, 1994
- [6] Tzung-Pei Hong and Chai-Ying Lee, "Induction of fuzzy rules and membership functions from training examples", *Fuzzy Sets and Systems* 84, pp. 33-47, 1996
- [7] A. Motro, "VAGUE: A User Interface to Relational Databases that Permit Vague Queries", *ACM trans. Office Information Systems*, Vol 6, No. 3, pp. 187-214, 1988
- [8] M. Zemankova and A. Kandel, "Implementing Imprecision in Information System", *Information Sciences*, Vol. 37, pp. 107-141, 1985