

정보검색에서 퍼지 언어 매트릭스에 근거한 효율적인 퍼지 질의 평가 방법

Effectual Fuzzy Query Evaluation Method based on Fuzzy Linguistic Matrix in Information Retrieval

최명복 · 김민구*

Myeong-Bok Choi and Min-Koo Kim

원주대학 행정전산과

* 아주대학교 컴퓨터공학과

요 약

본 논문에서는 시소러스에 근거한 새로운 퍼지 정보검색 기법을 제안한다. 제안된 방법에서 시소러스는 내부 용어들 간의 관련도를 정성적인 언어 값으로 갖는 퍼지 언어 매트릭스로 표현되며 용어들 간의 관계는 동의, 계층, 그리고 연관의 세 가지 관계가 제공된다. 시소러스 내부 용어들 간의 무시된 관련도가 퍼지 이론에 근거한 퍼지 언어 매트릭스의 전이폐쇄 알고리즘에 의해 추론된다. 또한 제안된 방법은 사용자의 질의, 그리고 문서와 같은 정보 항목의 표현에도 인간의 주관적이고 부정확한 측도를 그대로 반영하는 정성적인 언어 값을 허용한다. 따라서 논문[1-3]에서 제안된 방법보다 좀 더 유용하다. 또한 질의 평가시 퍼지 언어 매트릭스와 AON(Associated Ordinary Number)값을 이용하기 때문에 논문[1-3]에서 사용되는 방법보다 시간적으로 효율적이다. 결과적으로 사용자가 좀 더 유용하고 지능적인 방법으로 질의를 처리할 수 있도록 한다.

ABSTRACT

In this paper, we present a new fuzzy information retrieval method based on thesaurus. In the proposed method the thesaurus is represented by a fuzzy linguistic matrix, where the elements in the fuzzy linguistic matrix represent a qualitative linguistic values between terms. In the fuzzy linguistic matrix, there are three kinds of fuzzy relationships between terms, i.e., similar relation, hierarchical relation, and associative relation. The implicit fuzzy relationships between terms are inferred by the transitive closure of the fuzzy linguistic matrix based on fuzzy theory. And the proposed method has the capability to deal with a qualitative linguistic weights in a query and in indexing of information items to reflect qualitative measure of human based on vague and uncertain decisions rather than a quantitative measure. Therefore the proposed method is more flexible than the ones presented in papers[1-3]. Moreover our method is more effectual of time than the ones presented in papers[1-3] because we use a fuzzy linguistic matrix and AON(Associated Ordinary Number) values in query evaluation process. As a result, the proposed method allows the users to perform fuzzy queries in a more flexible and more intelligent manner.

1. 서 론

정보검색 시스템의 주요한 목적[4]을 극대화하기 위해서는 첫째 사용자의 정보요구와 문서의 색인인 검색형태(Search Patterns)를 명확하게 표현하는 것이고 [5], 둘째는 정보요구를 만족시키는 적절한 정보들만을 탐색하고 탐색된 정보들에 대해 정보요구의 만족도에 따른 적합성 순위를 부여하는 것이다. 대부분의 상업적인 정보검색 시스템들은 불리언 논리에 근거하고 있다. 그러나 이러한 시스템은 몇 가지 문제점[6]을 가지고 있다. 이와 같은 시스템에서 주요한 문제점 중의 하나는 주관성과 부정확성을 다루는 능력이

부족하다는데에 있다[7]. 이러한 시스템들은 인간의 주관적이고 부정확한 정보 표현을 허용하지 않기 때문에 부정확성이 존재하는 정보를 적절히 처리할 수 없다.

불리언 논리에 근거하는 검색 모델의 단점을 개선하기 위하여 시소러스와 같은 지식베이스를 이용하는 정보검색 방법들[2-3, 5, 8, 9-12]이 제안되었다. 그러나 이러한 방법들은 지식베이스 내부 용어들 간의 관계가 단일의 정량적인 숫자 값이며, 지식베이스의 추론 방법에 있어서 효율적이지 못하다.

논문[1]은 지식베이스에 근거하는 좀 더 효율적인 정보검색 방법을 제안하였다. 이 방법은 지식베이스를

구축하기 위하여 일반적으로 사용되는 그래프 형태의 개념 네트워크(Concept Network)[8]를 표현하기 위하여 개념 매트릭스(Concept Matrix)를 사용함으로써 효율적으로 질의를 평가한다. 또한 논문[1]의 방법은 사용자의 질의와 지식베이스 내부 용어들 간의 관련도를 정량적인 [0.0, 1.0] 사이의 구간 값(Interval Value)으로 부여할 수 있도록 하였다. 그러나 논문[1]의 방법에서 지식베이스 내부 용어들 간의 관계는 오직 한 가지의 퍼지 관계만을 표현할 수 있도록 하였다. 만약 지식베이스 내부 용어들 간에 여러 가지의 관련성을 제공하는 정보 검색 시스템이 있다면 좀 더 많은 유용성을 제공할 수 있을 것이다. 특히 사용자의 질의와 문서 표현, 그리고 지식베이스 내부 용어들 간의 관련도를 정성적으로 표현하는 정보 검색 시스템은 사용자에게 좀 더 유용한 검색 시스템의 역할을 제공한다고 볼 수 있다.

본 논문에서는 개념 매트릭스를 확장한 퍼지 언어 매트릭스를 이용하는 새로운 정보검색 방법을 제안한다. 퍼지 언어 매트릭스에서 용어들 간의 관계는 동의, 연관, 그리고 계층의 세 가지 관계로 구성된다. 퍼지 언어 매트릭스에서 용어들 간의 무시된 관련도가 퍼지 논리에 근거한 전이폐쇄 알고리즘에 의해 추론된다. 본 논문에서 제안된 방법은 사용자의 질의, 문서와 같은 정보 항목의 표현, 그리고 시소러스 내부 용어들 간의 관련도에 인간의 주관적이고 부정확한 측도를 그대로 반영하는 정성적인 언어 값을 제공하기 위해 퍼지 이론[13]에 근거한 언어적 설명자(Linguistic Descriptors)[13]를 이용하여 표현할 수 있도록 한다. 따라서 본 논문에서 제안된 방법은 논문[1-3]에서 제안된 방법보다 좀 더 유용하다. 또한 질의 평가시 퍼지 언어 매트릭스와 AON 값을 이용하기 때문에 논문[1-3]에서 사용되는 방법보다 시간적 효율성을 갖는다. 결과적으로 사용자가 좀 더 유용하고 지능적인 방법으로 질의를 처리할 수 있도록 한다.

2. 질의 표현을 위한 탐색어의 가중치 표현

질의의 탐색어에 부여하는 가중치는 여러 가지로 해석[14]되고 있지만 크게 두 가지 형태의 의미로 분류할 수 있다. 하나는 가중치를 그 탐색어의 중요성[7]으로 인식하며 다른 하나는 가중치를 한계 값(Threshold)[15]으로 받아드리는 것이다. 본 논문에서는 탐색어의 가중치를 전자의 의미인 탐색어의 중요성만을 고려한다.

산술 가중치는 질의와 정보항목의 검색 형태 표현

을 개선하기 위하여 사용되었다. 그러나 이러한 가중치의 부여는 사람의 주관적이고 부정확한 개념이다. 따라서 산술 가중치의 부여는 사람의 퍼지한 개념을 정확한 정량적 숫자 값으로 변환하기 위한 의미의 명확한 지식을 필요로 한다[7]. 이러한 문제를 해결하기 위하여 본 논문은 질의에 있는 탐색어에 정성적 가중치를 부여할 수 있도록 한다.

질의의 탐색어를 위한 정성적 가중치는 논문[16]에서 제안한 시소러스 내부 용어들 간의 관련도 표현과 유사하게 퍼지 이론[13]에 근거한 언어적 설명자[13]를 이용하여 다음과 같이 정의할 수 있다. 여기서 $T(I)$ 는 탐색어의 가중치를 나타내는 언어변수 I 의 항목집합이다.

$$T(I) = \{ \text{definitely important, extremely important, very important, important, slightly important, very slightly important, none important} \}$$

$T(I)$ 의 각 원소는 기본 항목 important에 대한 의미 규칙 M 의 정의에 의하여 그 의미가 부여된다. 탐색어의 가중치 부여는 시소러스 내부 용어들 간의 관련도의 부여와 밀접한 관련이 있다. 왜냐하면 시소러스 내부 용어들 간의 관련도는 탐색어에 부합되는 정보항목의 검색에 영향을 주기 때문이다. 따라서 본 논문

표 1. 가중치 표현에 대한 언어적 설명자와 퍼지 수의 대응
Table 1. Mapping between linguistic descriptor and fuzzy number for weights

언어적 설명자	TFN
definitely important(DI)	[1.000, 1.000, 1.000]
extremely important(EI)	[0.700, 0.850, 1.000]
very important(VI)	[0.525, 0.675, 0.825]
important(Imp)	[0.350, 0.500, 0.650]
slightly important(SI)	[0.175, 0.325, 0.475]
very slightly important(VSI)	[0.000, 0.150, 0.300]
none important(NI)	[0.000, 0.000, 0.000]

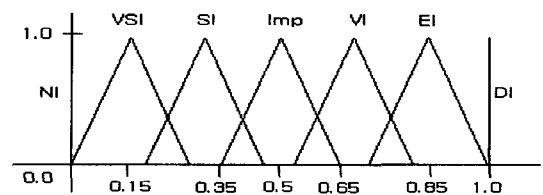


그림 1. 탐색어의 가중치에 대한 소속함수
Fig. 1. Membership function for weights of search term

서는 논문[16]에서 정의된 시소러스 내부 용어들 간의 관련도 부여와 일치하는 의미 표현을 제공하기 위하여 의미규칙 M 은 표 1, 그리고 그림 1과 같은 삼각형 소속 함수로 정의하였다.

3. 퍼지 언어 매트릭스

이 장에서는 효율적으로 질의를 평가하기 위해 개념 매트릭스를 확장한 퍼지 언어 매트릭스를 정의한다. 개념 매트릭스는 시소러스와 같은 그래프 형태의 지식베이스를 모델링 한다. 정보검색에서 일반적으로 시소러스는 노드와 링크로 구성된다. 노드는 용어를 링크는 용어들 간의 관계를 나타낸다. 논문[1]은 개념 매트릭스를 두 실수 구간으로 확장하여 용어들 간의 한 가지 관계에 적용하였다. 본 논문에서는 시소러스 내부 용어들 간에 언어적 관련도를 표현할 수 있도록 개념 매트릭스를 세 구간으로 확장하며 여러 가지 용어들 간의 관계에 적용한다.

논문[1]은 논문[8]에서 제시한 그래프 형태의 지식 베이스를 모델링 하기 위하여 다음과 같이 개념 매트릭스를 정의하였다.

[정의3.1] 개념들의 집합 $C = \{C_1, C_2, C_3, \dots, C_n\}$ 일 때 개념 매트릭스 M 은 퍼지 매트릭스[17]이다. 여기서 $M(C_i, C_j) \in [0, 1]$ 는 개념 C_i 로부터 C_j 까지의 관련도를 나타낸다. 개념 매트릭스 M 은 다음의 특성을 갖는다.

- 1) 반사관계
 $M(C_i, C_i) = 1, \forall C_i \in C.$
- 2) M 은 대칭관계가 아닐 수 있다.
 $M(C_i, C_j) \neq M(C_j, C_i).$
- 3) 전이관계
 $M(C_i, C_k) \geq \text{Max}_{C_j \in C} \text{Min}[M(C_i, C_j), M(C_j, C_k)].$

[정의3.2] 개념 매트릭스 M 은 다음과 같이 표현된다. 여기서 n 은 개념들의 개수이며, f_{ij} 는 개념 C_i 로부터 개념 C_j 까지의 관련 값(단, $f_{ij} \in [0, 1], 1 \leq i \leq n, 1 \leq j \leq n$)이다.

$$M = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1n} \\ f_{21} & f_{22} & \dots & f_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ f_{n1} & f_{n2} & \dots & f_{nn} \end{bmatrix}$$

논문[1]은 개념들 사이의 관련도를 0과 1 사이의 두 실수 구간으로 표현하기 위하여 다음과 같이 개념 매트릭스를 확장하였다.

[정의3.3] 두 구간에 대한 개념 매트릭스 M 은 다음과 같이 표현된다.

$$M = \begin{bmatrix} [f_{11(1)}, f_{11(h)}] & [f_{12(1)}, f_{12(h)}] & \dots & [f_{1n(1)}, f_{1n(h)}] \\ [f_{21(1)}, f_{21(h)}] & [f_{22(1)}, f_{22(h)}] & \dots & [f_{2n(1)}, f_{2n(h)}] \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ [f_{n1(1)}, f_{n1(h)}] & [f_{n2(1)}, f_{n2(h)}] & \dots & [f_{nn(1)}, f_{nn(h)}] \end{bmatrix}$$

여기서 n 은 개념들의 개수이며 $[f_{ij(0)}, f_{ij(h)}]$ (단, $0 \leq f_{ij(0)} \leq f_{ij(h)} \leq 1, i \leq n, 1 \leq j \leq n$)는 개념 C_i 로부터 C_j 까지 관련 값을 나타낸다.

이러한 개념 매트릭스는 논문[16]에서 제안된 시소러스 내부 용어들 간의 관련도를 정성적인 언어적 설명자로 표현할 수 있도록 세 구간으로 확장할 수 있다. 본 논문에서는 세 구간으로 확장된 매트릭스를 퍼지 언어 매트릭스(Fuzzy Linguistic Matrix)라고 명명할 것이다. 퍼지 언어 매트릭스는 다음과 같이 정의할 수 있다.

[정의 3.4] 용어들의 집합 $T_{or} = \{t_1, t_2, t_3, \dots, t_n\}$ 일 때 퍼지 언어 매트릭스 ML 은 구성 원소가 정성적인 언어 값을 갖는 퍼지 매트릭스[17]이다. 여기서 $ML(t_i, t_j)$ 은 인접한 용어 t_i 로부터 t_j 까지의 언어적 관련도를 의미한다.

언어적 관련도는 논문[16]에서 제안한 언어변수 E, H, R 의 각 항목집합 $T(E), T(H), T(R)$ 의 원소들 중 하나이다. 논문[16]에서는 용어들 간의 관계를 동의관계, 계층관계 그리고 연관관계의 3가지로 표현하고 이를 각각 E, H, R 의 약자로 표현하였다. 또한 E, H, R 을 언어변수로 하는 항목집합을 다음과 같이 퍼지 이론[13]에 근거한 언어적 설명자[13]를 이용하여 정의하였다.

- $T(E) = \{\text{completely similar(CS), very closely similar (VCS), very similar(VS), similar(Sim), slightly similar(SS), very slightly similar (VSS), none similar(NS)}\}$
- $T(H) = \{\text{definitely broader(DB), extremely broader (EB), very broader(VB), broader(Bro), slightly}$

표 2. 계층관계에 대한 언어적 설명자와 퍼지 수의 대응
Table 2. Mapping between linguistic descriptor and fuzzy number for $T(H)$

언어적 설명자	TFN
definitely broader(DB)	[1.000, 1.000, 1.000]
extremely broader(EB)	[0.700, 0.850, 1.000]
very broader(VB)	[0.525, 0.675, 0.825]
broader(Bro)	[0.350, 0.500, 0.650]
slightly broader(SB)	[0.175, 0.325, 0.475]
very slightly broader(VSB)	[0.000, 0.150, 0.300]
none broader(NB)	[0.000, 0.000, 0.000]

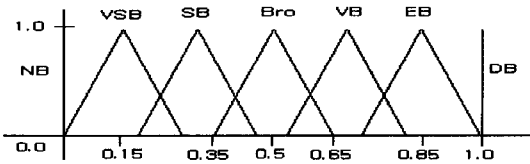


그림 2. 탐색어의 가중치에 대한 소속함수
Fig. 2. Membership function for $T(H)$

broader(SB), very slightly broader(VSB), none broader(NB)}

$T(R) = \{\text{certainly relevant(CR), very closely relevant (VCR), very relevant(VR), relevant(Rel), slightly relevant(SR), very slightly relevant (VSR), none relevant(NR)}\}$

그리고 $T(E)$, $T(H)$, 그리고 $T(R)$ 에 대한 소속함수의 정의는 $T(I)$ 와 같이 관련도의 등급에 따라 7개구간의 삼각형 퍼지 수로 분할된 퍼지 공간상에 표현하였다. 예를 들어 $T(H)$ 는 표 2, 그리고 그림 2와 같이 정의하였다.

퍼지 언어 매트릭스 ML 은 다음의 특성을 갖는다.

- 1) 반사관계
 $ML(t_i, t_i) = 1, \forall t_i \in T_{er}$.
- 2) 대칭관계
 $ML(t_i, t_j) = ML(t_j, t_i)$.
- 3) 전이관계가 아닐 수 있다(Nontransitive).
 $ML(t_i, t_k) \neq \text{Max}_{t_j \in T_{er}} \text{Min}[ML(t_i, t_j), ML(t_j, t_k)]$

[정의3.5] 세 구간에 대한 퍼지 언어 매트릭스 ML 은 다음과 같이 표현된다. 여기서 n 은 용어들의 개수이며 $[a_{ij(1)}, a_{ij(m)}, a_{ij(h)}]$ (단, $0 \leq a_{ij(1)}, a_{ij(m)}, a_{ij(h)} \leq 1, 1 \leq i \leq n, 1 \leq j \leq n$)는 삼각형 소속 함수로 인접한 용어 t_i 로부터 t_j 까지의 언어적 관련도를 나타낸다.

$$\begin{bmatrix} [a_{11(1)}, a_{11(m)}, a_{11(h)}] [a_{12(1)}, a_{12(m)}, a_{12(h)}] \\ [a_{21(1)}, a_{21(m)}, a_{21(h)}] [a_{22(1)}, a_{22(m)}, a_{22(h)}] \\ \vdots \vdots \\ [a_{n1(1)}, a_{n1(m)}, a_{n1(h)}] [a_{n2(1)}, a_{n2(m)}, a_{n2(h)}] \\ \dots [a_{1(1)}, a_{1n(m)}, a_{1n(h)}] \\ \dots [a_{2n(1)}, a_{2n(m)}, a_{2n(h)}] \\ \dots \vdots \\ \dots \vdots \\ \dots [a_{nn(1)}, a_{nn(m)}, a_{nn(h)}] \end{bmatrix}$$

4. 질의 평가 기법

정형적으로 정보검색 시스템은 $IRS = \langle T_{er}, Q_w, O_b, M_a \rangle$ 과 같이 4개의 구성요소를 갖는다. 여기서 T_{er} 는 용어들의 집합, Q_w 는 질의 집합, O_b 는 문서와 같은 정보 항목들을 표현하는 개체(Objects)의 집합, 그리고 M_a 은 질의와 개체 사이에 유사 정도를 계산하는 평가함수이다. 본 장에서는 질의 및 개체 표현 방법과 평가함수 M_a 에 대해 설명한다.

논문[1]과 비슷하게 질의 디스크립터 Q_w 는 질의 디스크립터 벡터 q 에 의해 다음과 같이 표현할 수 있다.

$$Q_w = \{(t_1, x_1), (t_2, x_2), \dots, (t_n, x_n)\}, q = \langle x_1, x_2, \dots, x_n \rangle, \text{(단, } x_i \in T(I), 1 \leq i \leq n) \quad (1)$$

여기서 x_i 는 탐색어 t_i 의 가중치로 2장에서 정의된 항목집합 $T(I)$ 의 원소중 하나이다. 질의 디스크립터 벡터 q 에서 $x_i=0$ 은 사용자가 용어 t_i 를 포함하지 않는 개체들을 검색하기를 요구한다는 의미이다. 또한, 사용자가 임의의 용어 t_i 를 질의에서 고려하지 않으려면 사용자는 가중치 x_i 를 명시하지 않으면 된다. 기호 “-”는 사용자에게 의해서 무시된 용어를 위해 사용될 수 있다. 예를 들면, 용어들의 집합 $T_{er} = \{t_1, t_2, t_3, t_4, t_5\}$ 이고, $Q_w = \{(t_2, VI), (t_4, EI), (t_5, NI)\}$ 이면 $q = \langle -, VI, -, EI, NI \rangle$ 과 같이 표현된다.

질의 디스크립터 Q_w 와 비슷하게 정보항목들을 표현하기 위한 O_b 는 개체 디스크립터로 표현할 수 있다. $O_b = \{o_1, o_2, \dots, o_n\}$ 이 정보검색 시스템의 데이터베이스를 구성하는 개체의 집합일 때 각각의 개체 o 는 다음의 식 (2)와 같이 색인어와 값의 쌍으로 구성되는 개체 벡터로 표현할 수 있다. 여기서 n 은 용어 집합 T_{er} 의 원소 수이며 w 는 그 개체에 대한 색인어

t_i 의 가중치를 나타낸다. 본 논문은 이러한 가중치의 의미를 질의의 탐색어에 주어지는 가중치의 의미와 동일하게 해석한다. 따라서 본 논문에서는 개체에 대한 색인어의 가중치를 2장에서 정의된 언어변수 I 의 항목집합 $T(I)$ 의 원소들로 부여한다. 예를 들어 $o = \langle (t_1, VI), (t_2, ED) \rangle$ 일 때 개체 o 에 대한 t_1 과 t_2 의 가중치는 각각 VI, ED 가 된다. 이러한 의미는 t_1 과 t_2 가 개체 o 에 각각 VI, ED 의 관련도로 관계되어 있다고 볼 수 있다.

$$o = \langle (t_1, w_1), (t_2, w_2), \dots, (t_m, w_m) \rangle \quad (2)$$

(단, $t_i \in T_{o_i}, w_i \in T(I), 1 \leq i \leq m$)

개체 집합 O_b 는 가중치 w 에 대해 o 와 i 간의 개체 디스크립터 벡터로 다음과 같이 표현할 수 있다. 여기서 m 은 개체들의 수이고 n 은 용어들의 수이다. w_{ij} (단, $w_{ij} \in T(I), 1 \leq i \leq m, 1 \leq j \leq n$)는 개체 o_i 에 대한 용어 t_j 의 가중치를 나타낸다.

$$O_{bj} = \begin{matrix} & t_1 & t_2 & t_3 & \dots & t_n \\ \begin{matrix} o_1 \\ o_2 \\ o_3 \\ \vdots \\ \vdots \\ o_M \end{matrix} & \begin{bmatrix} w_{11} & w_{12} & w_{13} & \dots & w_{1n} \\ w_{21} & w_{22} & w_{23} & \dots & w_{2n} \\ w_{31} & w_{32} & w_{33} & \dots & w_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ w_{m1} & w_{m2} & w_{m3} & \dots & w_{mn} \end{bmatrix} \end{matrix}$$

개체 디스크립터 매트릭스 O_{bj} 에서 개체 o_i 에 대한 용어 t_j 의 관련성 정도는 전문가에 의해 결정될 수 있다. 그러나 전문가는 어떤 용어들에 대해서는 특정 개체로의 관련성을 무시할 수 있다[1]. 무시된 관련성은 용어와 용어사이의 관련성을 나타내는 퍼지 언어 매트릭스의 모든 연결 관계를 찾아봄으로써 구할 수 있다. 이러한 퍼지 언어 매트릭스에서 용어들 간의 모든 연결 관계는 퍼지 언어 매트릭스 ML 의 전이폐쇄 (Transitive Closure)에 의해 쉽게 파악할 수 있다. 퍼지 언어 매트릭스 ML 의 전이폐쇄 $T(ML)$ 은 다음과 같은 간단한 알고리즘[18]에 의해 구할 수 있다.

[정의4.1] 관계 $R(X, X)$ 가 있을 때, 관계 $R(X, X)$ 의 전이폐쇄 $R_T(X, X)$ 는 다음과 같은 3단계로 구성되는 간단한 알고리즘에 의해 결정될 수 있다.

1. $R' = R \cup (R \circ R)$.
2. If $R' \neq R$, make $R = R'$ and go to Step 1.
3. Stop: $R' = R_T$

전이폐쇄 알고리즘의 단계 1에 있는 합집합 기호 \cup 와 합성 기호 \circ 는 여러 가지의 연산자가 적용될 수 있다. 본 논문에서는 퍼지 관계 $R(X, X)$ 에 대한 다음과 같은 두 가지의 연산자를 전이폐쇄 알고리즘에 적용할 것이다.

$$\mu_R(x, z) = \text{Max}_{y \in Y} \text{Min}[\mu_R(x, y), \mu_R(y, z)] \quad (3)$$

(단, $(x, z) \in X^2$)

$$\mu_R(x, z) = \text{Max}_{y \in Y} [\mu_R(x, y) \cdot \mu_R(y, z)] \quad (4)$$

(단, $(x, z) \in X^2$)

수식 (3), (4)의 최대-최소 합성, 최대-곱 합성은 논문[16]에서 제안된 인식론적 관점에 근거한 추론 기법(다음의 수식 (5)-(8)의 의미를 그대로 반영한다. 수식에서 t_i, t_j, t_k 는 용어이며, $D_{i,j}, D_{j,k}$ 은 각각 t_i 와 t_j, t_j 와 t_k 사이의 언어적 관련도로 $T(E), T(H), T(R)$ 의 원소로 대신할 수 있다. $D_{i,k}(E), D_{i,k}(H), D_{i,k}(R)$ 은 $D_{i,j}, D_{j,k}$ 을 이용하여 t_i 와 t_k 간에 추론된 관련도로 각각 동의, 계층, 연관관계로 추론된다는 것을 표시한다. 관련도의 연산은 다음 장에서 상세히 다룬다.

$$t_i = t_k \text{ 경우,} \\ D_{i,k}(E) = D_{i,k}(H) = D_{i,k}(R) = 1 \quad (5)$$

$$t_i \neq t_k \text{ 경우,} \\ D_{i,k}(E) = \text{Max}_j \text{Min}[D_{i,j}, D_{j,k}] \quad (6)$$

$$D_{i,k}(H) = \text{Max}_j (D_{i,j} \cdot D_{j,k}) \quad (7)$$

$$D_{i,k}(R) = \text{Max}_j (D_{i,j} \cdot D_{j,k}) \quad (8)$$

이상에서 설명된 퍼지 언어 매트릭스 ML 의 전이폐쇄 $T(ML)$ 를 개체 디스크립터 매트릭스 O_{bj} 에 적용하면 o_i 와 t_j 사이에 명시되지 않은 모든 관련도를 추론해 내게 된다. 관련도 $O_{bj}(*)$ 는 다음의 수식 (9)과 같이 구할 수 있다. 여기서 $O_{bj}(*)$ 는 특정 용어에 대한 각 개체의 관련성 정도를 나타내며, 뒤에서 질의와 각 개체 사이의 유사 측도에 사용할 것이다. 수식 (9)은 아래의 [정의4.2]에 근거하여 식 (10)과 같이 연산한다.

$$O_{bj}(*) = O_{bj} \odot T(ML) \quad (9)$$

수식 (10)에서 \hat{w}_{ij} 와 \hat{D}_{ij} 는 w_{ij}, D_{ij} 가 $[0.0, 1.0]$ 사이의 세 값으로 구성되는 삼각형 퍼지 수로 각각 개체 디스크립터 매트릭스 O_{bj} 와 퍼지 언어 매트릭스 ML

의 전이폐쇄 $T(ML)$ 의 원소일 때, w_{ij}, D_{ij} 에 대한 AON(Associated Ordinary Number) 값이다. AON 값은 다음 장에서 상세히 설명된다.

[정의4.2] [정의3.4]에서와 같은 특성을 갖는 퍼지 언어 매트릭스 ML 이 있을 때, O_{ij} 와 $T(ML)$ 의 최대-최소 합성(Max-Min)은 o_i 에 대한 각 t_j 의 모든 연결 관계를 표시한다.

$$O_{ij} = \begin{matrix} & t_1 & t_2 & \dots & t_n \\ \begin{matrix} O_1 \\ O_2 \\ \vdots \\ O_m \end{matrix} & \begin{matrix} v(\hat{w}_{1i} \vee \hat{D}_{i1}) \\ v(\hat{w}_{2i} \vee \hat{D}_{i1}) \\ \vdots \\ v(\hat{w}_{mi} \vee \hat{D}_{i1}) \end{matrix} & \begin{matrix} v(\hat{w}_{1i} \vee \hat{D}_{i2}) \\ v(\hat{w}_{2i} \vee \hat{D}_{i2}) \\ \vdots \\ v(\hat{w}_{mi} \vee \hat{D}_{i2}) \end{matrix} & \dots & \begin{matrix} v(\hat{w}_{1i} \vee \hat{D}_{in}) \\ v(\hat{w}_{2i} \vee \hat{D}_{in}) \\ \vdots \\ v(\hat{w}_{mi} \vee \hat{D}_{in}) \end{matrix} \end{matrix} \quad (10)$$

지금부터는 이상에서 설명된 정보검색 시스템의 구성요소인 T_{er}, Q_{er}, O_{ij} 를 이용하여 평가 함수 M_{ar} 을 구할 것이다. 평가 함수 M_{ar} 은 개체집합 O_{ij} 에서 질의 q 에 적합한 정보항목들을 찾아내는 함수이다. 본 논문에서 질의에 있는 탐색어의 가중치는 개체 디스크립터 매트릭스 O_{ij} 에 있는 각 개체들을 위한 색인어들의 가중치와 같은 의미의 가중치로 해석한다. 따라서 탐색어의 가중치와 같은 정도로 부여된 색인어의 가중치는 사용자가 찾고자 하는 개체에 대한 가중치의 의도와 완전히 부합한다. 이러한 가중치의 의미 관계를 반영하기 위하여 평가 함수 M_{ar} 은 두 실수간의 유사 측도 방법을 적용할 수 있다. 본 논문에서는 논문 [19,20]에서 제안된 유사 측도 방법을 삼각형 소속함수로 확장하여 적용한다. 두 삼각형 퍼지 수 A, B간의 유사 정도는 함수 S에 의해 다음과 같이 계산할 수 있다.

$$S(A, B) = 1 - |\hat{A} - \hat{B}| \quad (\text{단, } A = [a_1, a_2, a_3], B = [b_1, b_2, b_3]) \quad (11)$$

수식에서 $S(A, B)$ 의 값이 크면 클수록 A와 B사이의 유사 정도는 높아지게 된다. 수식 (11)에 있는 삼각형 퍼지 수에 대한 \hat{A}, \hat{B} 값은 AON 값으로 다음 장에서 상세히 설명된다.

개체 디스크립터 벡터 o_i 와 질의 디스크립터 벡터 q 가 다음과 같을 때 o_i 와 q 사이의 유사 정도는 수식

(11)을 이용하여 수식 (12)과 같이 RSV(Retrieval Status Values)로 계산할 수 있다. 여기서 n 은 용어들의 개수, m 은 개체들의 개수이다.

$$o_i = \langle w_{i1}, w_{i2}, \dots, w_{in} \rangle, q = \langle x_1, x_2, \dots, x_n \rangle$$

(단, $w_{ij} \in T(I), x_j \in T(I), 1 \leq j \leq n, 1 \leq i \leq m$)

$$RSV(o_i) = \frac{\sum S(w_{ij}, x_j) \quad \text{단, } q(j) \neq "-" \text{ and } j=1, 2, \dots, n}{k}$$

(단, $0 \leq RSV(o_i) \leq 1$) (12)

여기서 $q(j)$ 는 질의 디스크립터 벡터 q 의 j 번째 요소를 의미하며 k 는 질의 디스크립터 벡터에 명시된 용어들의 개수이다. $q(j) = "-"$ 은 용어 t_j 가 무시되었음을 의미한다. 이와 같이 명시되지 않은 용어는 검색 과정에서 불필요하기 때문에 연산에서 제외된다.

3장에서 시소러스 내부 용어들 간에는 세 가지 관계(동의, 계층, 연관)가 제공되었다. 관계를 구별하는 이유는 용어들 간의 추론에서 두 용어들이 동의 관계로만 구성된 경우, 계층 관계로만 구성된 경우, 연관 관계로만 구성된 경우, 그리고 이들 세 관계들이 혼합된 경우 서로 다른 추론 결과가 계산되어 질의 평가에 반영되도록 하는데 있다. 관계들 각각의 추론 결과는 [정의4.1]에서 이미 반영하였다. 따라서 이상의 수식 (9)-(12)에서는 세 가지 관계들을 구별하지 않고 AON 값만을 이용하게 된다.

질의는 AND 또는 OR로 구성되는 두 가지 형태로 분류할 수 있다. AND로 연결된 질의는 탐색어들의 AND 연결을 나타내는 질의 디스크립터 벡터에 의해 쉽게 표현된다.

$$q = \langle x_1, x_2, \dots, x_n \rangle \quad (\text{단, } x_j \in T(I), 1 \leq j \leq n) \quad (13)$$

예를 들어 $q = \langle _, VI, _, _, EI, DI \rangle$ 는 용어 t_2 가 VI이고 t_4 가 EI, 그리고 t_5 가 DI의 가중치를 갖는 개체들을 검색하기를 원하는 것으로 볼 수 있다.

또한 OR로 연결된 질의는 다음의 표현처럼 AND로 연결된 질의 디스크립터들을 OR로 연결한 질의를 의미한다.

$$q_1 \text{ OR } q_2 \quad (14)$$

예를 들어 $q_1 = \langle _, VI, _, _, _ \rangle$ OR $q_2 = \langle _, _, SI, _, _ \rangle$ 는 용어 t_2 가 VI의 가중치 갖고 있든지 또는 t_3 가 SI의 가중치를 가지고 있는 개체들을 검색하고자 하는 질의로 볼 수 있다. 따라서 OR로 연결된 질의와 개체 사이의 유사 정도는 다음과 같이 계산할 수

있다. 수식에서 m 은 개체의 개수이다.

$$RSV^*(o_i) = \text{Max}[RSV_1(o_i), RSV_2(o_i)] \quad (15)$$

(단, $0 \leq RSV^*(o_i) \leq 1, 1 \leq i \leq m$)

여기서 RSV_j 는 질의 디스크립터 벡터 내에 있는 각각의 q_j 와 개체 디스크립터 벡터의 각 행에 있는 o_i 사이의 유사 정도이다. 따라서 $RSV_1(o_i)$ 는 질의 디스크립터 벡터 q_1 과 개체 디스크립터 매트릭스의 i 행에 있는 o_i 사이의 유사 정도이며 $RSV_2(o_i)$ 는 질의 디스크립터 벡터 q_2 와 개체 디스크립터 매트릭스의 i 행에 있는 o_i 사이의 유사 정도이다. 검색 상태 값 $RSV^*(o_i)$ 는 질의와 개체 o_i 사이의 유사 정도를 나타낸다.

5. 삼각형 소속함수의 연산 방법

본 장에서는 4장의 수식 (3)-(8)에서 사용되는 삼각형 퍼지 수들 간의 연산자들을 정의한다. 또한 수식 (10)-(11)을 위한 AON값의 연산과 AON값에 근거한 추론된 관련도의 언어 근사(Linguistic Approximation)[13] 및 분류(Classification) 방법을 설명한다.

그림 1과 같이 제안된 삼각형 퍼지 수는 전체집합 $U=[0.0, 1.0]$ 을 갖는 퍼지 공간을 논문[16]의 [가정2]에 근거하여 일정한 비율로 분할하였다. 이러한 분할은 전체집합 상에서 선형순서(Linear Ordering)[21]를 갖게 한다. 따라서 제안된 삼각형 퍼지 수들은 전체집합 상의 수평이동으로 볼 수 있다.

논문[22,23]은 인간의 언어적 직관을 좀 더 잘 반영하기 위하여 소속함수의 이동(Shifting)이동에 대해 다음과 같은 수식을 제안하였다.

$$M(\mathcal{M}A) = \nu\mathcal{M} \circ A \circ \xi\mathcal{M} \quad (16)$$

(단, $\nu\mathcal{M} : [0, 1] \rightarrow [0, 1], \xi\mathcal{M} : U \rightarrow U$)

수식에서 A 는 기본 함수, $\nu\mathcal{M}$ 은 변형함수(Modification Function), 그리고 $\xi\mathcal{M}$ 은 이동함수(Shifting Function)를 나타낸다. 따라서 제안된 항목집합 $T(E), T(H), T(R)$, 그리고 $T(I)$ 의 원소들에 대한 삼각형 퍼지 수들은 수식 (16)에 용이하게 적용할 수 있다. 예를 들어 그림 1에서 E 는 V 로부터 0.175 만큼 오른쪽으로 평행 이동한 삼각형 퍼지 수로 볼 수 있다. 이 경우 $\nu\mathcal{M}$ 은 무시되고 A 는 V , $\xi\mathcal{M}$ 은 0.175이며, 연산자 \circ 는 +로 해석할 수 있다.

수식 (16)에서 합성 연산자 \circ 는 문제 영역에 따라 여러 가지가 적용될 수 있다. 민코스키(Minkoski)의 연산자들[21]은 수식 (16)에서 A 에 대한 이동함수 $\xi\mathcal{M}$ 을 용이하게 모델링 한다. 따라서 본 논문에서는 두 삼각형 퍼지 수의 연산을 위해 민코스키의 연산

방법을 사용할 것이다. 민코스키의 연산은 두 언어적 라벨이 표 1에서 정의된 삼각형 퍼지 수와 같이 $A_1 = [a_1, a_2, a_3]$ 와 $A_2 = [b_1, b_2, b_3]$ 와 같이 표현되었을 때 다음과 같은 몇 가지 연산이 정의된다[21].

[정의5.1] 덧셈 : $A_1(+)A_2 = [a_1, a_2, a_3] (+) [b_1, b_2, b_3]$

$$m = [a_1 + b_1, a_2 + b_2, a_3 + b_3]$$

곱셈 : $A_1(\cdot)A_2 = [a_1, a_2, a_3] (\cdot) [b_1, b_2, b_3]$

$$m = [a_1 \cdot b_1, a_2 \cdot b_2, a_3 \cdot b_3]$$

이상에서 정의된 연산자에 추가하여 몇 가지 연산 수식이 더 필요하다. 여기서 k 는 상수이며 \wedge, \vee 는 각각 최소값(Minimum)과 최대값(Maximum)을 의미한다.

[정의5.2] 상수 나눗셈 : $A/k = [a_1, a_2, a_3]/k$

$$= [a_1/k, a_2/k, a_3/k]$$

최소값 : $\text{Min}(A_1, A_2) = [a_1, a_2, a_3] (\wedge) [b_1, b_2, b_3]$

$$= [a_1 \wedge b_1, a_2 \wedge b_2, a_3 \wedge b_3]$$

최대값 : $\text{Max}(A_1, A_2) = [a_1, a_2, a_3] (\vee) [b_1, b_2, b_3]$

$$= [a_1 \vee b_1, a_2 \vee b_2, a_3 \vee b_3]$$

[정의5.2]에 있는 최소값, 최대값은 삼각형 퍼지 수들 간의 일반적인 연산자[21]이다. 또한 [정의5.1]과 [정의5.2]에 있는 연산자들은 수식 (16)의 이동함수를 용이하게 반영한다. 이상에서 정의된 삼각형 퍼지 수들 간의 연산자들을 수식 (3)-(8)에 적용하면 퍼지 언어 매트릭스 ML 의 전이패쇄 $T(ML)$ 을 얻을 수 있다.

$T(ML)$ 을 이용하여 수식 (10)의 $O_b(*)$ 를 계산하기 위하여 AON(Associated Ordinary Number)에 대한 정의가 필요하다. AON[21] 값은 $A = [a_1, a_2, a_3]$ 와 같은 삼각형 퍼지 수에 대해 다음과 같은 등식으로 표현되는 선형 순서 방법이다.

$$\hat{A} = (a_1 + 2(a_2) + a_3)/4 \text{ (단, } A = [a_1, a_2, a_3]) \quad (17)$$

식 (17)을 제안된 유사 측도 수식 (11)에 적용하여 수식 (12)을 이용하면 질의의 각 정보항목들에 대해 검색 상태 값(RSV)이 계산된다. 계산된 RSV는 수식 (11)에서 AON 값을 이용하기 때문에 $U = [0.0, 1.0]$ 상의 단일 값이다. 예를 들어 계산된 RSV는 아래의 그림 3에서 2개의 굵은 실선으로 되어 있는 삼각형 퍼지 수와 같이 표현되지 않고 VSR과 SR 또는 Rel과 VR 사이의 수직선과 같이 단일 값으로 퍼지 공간 상에 위치한다. 이러한 단일 값들은 삼각형 퍼지 수에 대한 AON 값으로 볼 수 있으며 RSV의 순서화는 단지 그 계산된 RSV 값들을 크기순으로 나열만 하면 된다.

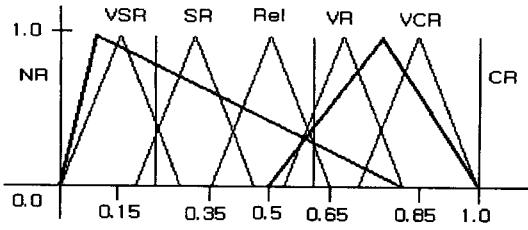


그림 3. 삼각형 퍼지 수의 언어근사 및 분류
Fig. 3. Linguistic approximation and classification of TFN

이와 같이 AON에 근거한 RSV의 순서화는 언어 근사화 및 분류를 용이하게 한다. 언어적 분류는 질의에 대한 검색된 정보항목들이 많은 경우 모든 정보항목들을 훑어 보아야 하는 사용자의 노력을 감소시킬 수 있으며, 검색된 정보항목들을 선택하는데 있어서 수량적 RSV 보다는 좀 더 강력하게 관련성을 제시해 줄 수 있다[7]. 본 논문에서는 이미 설명된 것처럼 탐색어와 색인어의 가중치, 그리고 시소러스 내부 용어들 간의 관련도 부여를 위한 언어적 설명자들에 대한 의미가 이동함수에 근거한 소속함수로 정의되었기 때문에 언어 근사 및 분류를 위한 새로운 언어적 라벨을 필요로 하지 않는다. 본 논문에서는 RSV에 대한 언어근사 및 분류를 위해 용어들 간의 관련도에 사용된 $T(R)$ 를 사용할 것이다. $T(R)$ 의 사용은 RSV 값이 이동함수에 근거하여 정의된 언어적 설명자를 이용하여 추론되었기 때문에 자연스럽다. 예를 들어 그림 3에서 VSR과 SR 또는 Rel과 VR 사이의 수직선의 AON 값이 각각 $RSV_1=0.237$, $RSV_2=0.590$ 인 경우를 생각해 보자. 우선, NR(none relevant), VSR (very slightly relevant), SR(slightly relevant), Rel (relevant), VR(very relevant), VCR(very closely relevant), CR(certainly relevant) 각각을 수식 (17)에 적용하면 각각 0.0, 0.15, 0.325, 0.5, 0.675, 0.85, 1.0의 AON 값을 얻게 된다. 두 AON 값 0.237, 0.590은 각각 VSR, VR의 AON 값에 가장 근사해 있다. 그러므로 RSV_1 , RSV_2 는 각각 VSR과 VR로 언어 근사 및 분류된다.

이제 다음과 같은 [예제 5.1]을 가지고 사용자 질의에 대한 각 정보항목의 검색 상태 값(RSV)을 계산해 보자.

[예제 5.1] 아래의 예제는 7개의 정보항목 o_1, o_2, \dots, o_7 과 7개의 용어 t_1, t_2, \dots, t_7 를 가정한다. 또한 개체 디스크립터 매트릭스 O_b 와 퍼지 언어 매트릭스 ML 은 다음과 같이 구성되어 있으며, 퍼지 언어 매트릭스 ML 에서 용어들 간의 관계는 동의관계로만 구성됨을 가정한다.

$$O_{bj} = \begin{bmatrix} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 \\ \text{Imp} & \text{VI} & \text{DI} & \text{NI} & \text{NI} & \text{VI} & \text{NI} \\ \text{DI} & \text{VI} & \text{NI} & \text{SI} & \text{DI} & \text{NI} & \text{NI} \\ \text{NI} & \text{DI} & \text{NI} & \text{Imp} & \text{Imp} & \text{SI} & \text{DI} \\ \text{VI} & \text{Imp} & \text{DI} & \text{SI} & \text{NI} & \text{VI} & \text{DI} \\ \text{DI} & \text{NI} & \text{VI} & \text{DI} & \text{NI} & \text{Imp} & \text{VI} \\ \text{EI} & \text{SI} & \text{Imp} & \text{VI} & \text{DI} & \text{NI} & \text{DI} \\ \text{NI} & \text{EI} & \text{EI} & \text{EI} & \text{NI} & \text{DI} & \text{DI} \end{bmatrix}$$

$$ML = \begin{bmatrix} o_1 & t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 \\ \text{CS} & \text{CS} & \text{NS} & \text{NS} & \text{NS} & \text{NS} & \text{NS} & \text{NS} \\ o_2 & \text{NS} & \text{CS} & \text{SS} & \text{NS} & \text{NS} & \text{NS} & \text{VCS} \\ o_3 & \text{NS} & \text{SS} & \text{CS} & \text{NS} & \text{NS} & \text{NS} & \text{Sim} \\ o_4 & \text{NS} & \text{NS} & \text{NS} & \text{CS} & \text{CS} & \text{CS} & \text{NS} \\ o_5 & \text{NS} & \text{NS} & \text{NS} & \text{NS} & \text{CS} & \text{NS} & \text{VCS} \\ o_6 & \text{NS} & \text{NS} & \text{NS} & \text{NS} & \text{NS} & \text{CS} & \text{VS} \\ o_7 & \text{NS} & \text{VCS} & \text{Sim} & \text{NS} & \text{VCS} & \text{VS} & \text{CS} \end{bmatrix}$$

이와 같은 경우 퍼지 언어 매트릭스 ML 에 대한 전이폐쇄 $T(ML)$ 는 [정의 4.1]에 의해 다음과 같이 구해진다.

$$T(ML) = \begin{bmatrix} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 \\ t_1 & \text{CS} & \text{CS} & \text{CS} & \text{NS} & \text{VCS} & \text{VS} & \text{VCS} \\ t_2 & \text{NS} & \text{CS} & \text{Sim} & \text{NS} & \text{VCS} & \text{VS} & \text{VCS} \\ t_3 & \text{NS} & \text{Sim} & \text{CS} & \text{NS} & \text{Sim} & \text{Sim} & \text{Sim} \\ t_4 & \text{NS} & \text{VCS} & \text{Sim} & \text{CS} & \text{CS} & \text{CS} & \text{VCS} \\ t_5 & \text{NS} & \text{VCS} & \text{Sim} & \text{NS} & \text{CS} & \text{VS} & \text{VCS} \\ t_6 & \text{NS} & \text{VS} & \text{Sim} & \text{NS} & \text{VS} & \text{CS} & \text{VS} \\ t_7 & \text{NS} & \text{VCS} & \text{Sim} & \text{NS} & \text{VCS} & \text{VS} & \text{CS} \end{bmatrix}$$

개체 디스크립터 매트릭스 $O_b(*)$ 는 개체 디스크립터 매트릭스 O_b 와 퍼지 언어 매트릭스 ML 의 전이폐쇄 $T(ML)$ 를 이용하여 다음과 같이 AON 값으로 구해진다.

$$O_{bj}(*) = \begin{bmatrix} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 \\ o_1 & 0.500 & 0.675 & 1.000 & 0.000 & 0.675 & 0.675 & 0.675 \\ o_2 & 1.000 & 1.000 & 1.000 & 0.325 & 1.000 & 0.675 & 0.850 \\ o_3 & 0.000 & 1.000 & 0.500 & 0.500 & 0.850 & 0.675 & 1.000 \\ o_4 & 0.675 & 0.850 & 1.000 & 0.325 & 0.850 & 0.675 & 1.000 \\ o_5 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 & 0.850 \\ o_6 & 0.850 & 0.850 & 0.850 & 0.675 & 1.000 & 0.675 & 1.000 \\ o_7 & 0.000 & 0.850 & 0.850 & 0.850 & 0.850 & 1.000 & 1.000 \end{bmatrix}$$

만약 사용자의 질의가 다음과 같이 OR로 연결된 경우 각 정보항목에 대한 검색 상태 값(RSV)은 수식 (12)에 의해 다음과 같이 얻어진다.

Query : $\langle VI, _ _ Imp, EI, _ _ \rangle$ OR $\langle _ _ EI, Imp, _ _ _ \rangle$

- $RSV_1(o_1) = 0.717, RSV_2(o_1) = 0.662$
- $RSV_1(o_2) = 0.783, RSV_2(o_2) = 0.675$
- $RSV_1(o_3) = 0.775, RSV_2(o_3) = 0.925$
- $RSV_1(o_4) = 0.942, RSV_2(o_4) = 0.750$
- $RSV_1(o_5) = 0.675, RSV_2(o_5) = 0.675$
- $RSV_1(o_6) = 0.833, RSV_2(o_6) = 0.825$
- $RSV_1(o_7) = 0.658, RSV_2(o_7) = 0.825$

위의 결과에 수식 (15)을 적용하면 다음과 같은 결과를 얻는다.

- $RSV^*(o_1) = 0.717$
- $RSV^*(o_2) = 0.783$
- $RSV^*(o_3) = 0.925$
- $RSV^*(o_4) = 0.942$
- $RSV^*(o_5) = 0.675$
- $RSV^*(o_6) = 0.833$
- $RSV^*(o_7) = 0.825$

따라서 앞에서 설명된 것과 같이 AON 값에 근저한 RSV의 순서화는 다음과 같이 언어근사 및 분류된 결과를 얻게 된다.

- $RSV^*(o_4) = 0.942$ (CR)
- $RSV^*(o_3) = 0.925$ (CR, VCR)
- $RSV^*(o_2) = 0.783$ (VCR)
- $RSV^*(o_6) = 0.833$ (VCR)
- $RSV^*(o_7) = 0.825$ (VCR)
- $RSV^*(o_1) = 0.717$ (VR)
- $RSV^*(o_5) = 0.675$ (VR)

위의 결과로부터 정보항목 o_4 는 가장 큰 검색 상태 값을 갖기 때문에 사용자의 질의에 가장 적합하다. 또한 각각의 정보항목은 괄호 내의 언어 값으로 근사되며 CR, VCR, VR의 세 그룹으로 분류된다.

6. 결 론

본 논문에서는 사용자의 질의, 문서와 같은 정보 항목의 표현, 그리고 용어들 간의 관련도에 인간의 주관적이고 부정확한 측도를 그대로 반영하는 정성적인 언어 값을 허용하며 시소러스에서 용어들 간에 여러 가지 관련성을 제공하였다. 따라서 논문[1-3]에서 제안된 방법보다 좀 더 유용하다.

논문[2-3]에서는 질의를 평가하기 위해 그래프 형태의 지식베이스를 이용하기 때문에 n 개의 정보항목에 대해 지식베이스를 최소 n 번 이상 실시간으로 탐색해

야 한다. 특히 실제적인 시스템에서 사용자의 질의가 복합 질의(AND나 OR 등의 연결자를 사용하여 구성된 질의)로 구성되고 지식베이스 내부 용어들 간의 관계가 여러 경로를 통해 연결된 다중 경로로 구성되어 있는 경우 그 추론 절차는 더욱 느리게 되며 비효율적이다. 그러나 본 논문에서는 질의 평가시 논문[16]에서 제안된 인식론적 추론기법을 용이하게 모델링 하는 퍼지 언어 매트릭스를 정의하여 효율적인 질의 평가가 이루어지도록 하였다. 또한 논문[1]에서는 질의 평가시 두 구간 연산을 수행하기 때문에 본 논문에서 제안된 AON 값을 이용하는 방법보다는 덜 효율적이다. 결과적으로 본 논문에서 제안된 방법은 사용자에게 좀 더 유용하고 지능적인 방법으로 질의를 처리할 수 있도록 하였다.

논문[1]은 컴퓨터 과학에 관련된 100권의 문서와 19개의 색인어를 사용하여 질의를 평가하였다. 실험에서 탐색어 및 색인어의 가중치와 지식베이스 내부 용어들 사이의 관련도가 두 구간 값으로 주어졌다. 실험 결과에서 탐색어의 가중치를 정량적인 [0.0, 1.0] 내의 단일 값으로 부여했을 때와 두 구간 값으로 부여했을 때의 정확율(Precision Rate)[6]은 일치했지만 구간 값으로 부여하는 것이 좀 더 사용자에게 유용하다고 결론지었다. 본 논문에서 제안된 방법에서 정성적으로 가중치 및 관련도를 부여하며 지식베이스 내부 용어들 간의 관계를 동의관계로 제한한다면, 논문[1]의 실험과 같은 문서 및 지식베이스 환경에서는 같은 결과가 얻어질 것이다. 이것은 특히 지식베이스 내부 용어들 간의 추론 방법으로 두 방법 모두 최대-최소(Max-Min)합성을 이용하기 때문이다.

제안된 모델은 정보항목을 검색하는 검색 시스템에 쉽게 적용될 수 있다. 예를 들어 정보항목이 문서인 문서 검색 시스템 또는 정보항목이 원본코드(Source Codes)인 소프트웨어 재사용 라이브러리 시스템(Software Reuse Library System)[2]등이 있겠다.

앞으로의 연구 방향은 용어들 간의 여러 가지 관계를 표현하는 지식베이스를 구축하여 사용자의 질의를 다양하게 평가해야 할 것이다.

참고문헌

- [1] Chen, S. M.; Wang, J. Y, Document retrieval using knowledge-based fuzzy information retrieval techniques, *IEEE Transactions on systems, man, cybernetics*, Vol. 25, No. 5, pp. 793-803, 1995.
- [2] Kim, Y. W.; Kim, J. H. A model of knowledge based information retrieval with hierarchical concept graph, *Journal of Documentation*, 46(2), pp. 113-136, 1990.

- [3] Rada, R.; Mili, H.; Bicknell, E., & Blettner, M. Development and application of a metric on semantic nets, *IEEE Transactions on systems, man, cybernetics*, Vol. 19, No. 1, pp. 17-36, January, 1989.
- [4] Lee, J. H., Kim, M. H., & Lee, J. H. Ranking documents in thesaurus-based boolean retrieval systems, *Information Processing & System*. 30(1), pp. 79-91, 1994.
- [5] Radecki, T. Fuzzy set theoretical approach to document retrieval, *Information Processing & Management*, Vol. 15, pp. 247-259, 1979.
- [6] Salton, G.; McGill, M. J. Introduction to modern information retrieval, New York: McGraw-Hill, 1983.
- [7] Bordogna, G.; Pasi, G. A fuzzy linguistic approach generalizing boolean information retrieval: A model and its evaluation, *Journal of the American society for information science*, 44, pp. 70-82, 1993.
- [8] Lucarella, D.; Morara, R. FIRST: Fuzzy information retrieval system, *Journal of Information Science*, Vol. 17, pp. 81-91, 1991.
- [9] McMath, C. F.; Tamaru, R. S.; Rada, R. A graphical thesaurus-based information retrieval system, *International Journal of Man-Machine Studies*, 31(2), pp. 121-147, 1989.
- [10] Radecki, T. Mathematical model of information retrieval system based on the concept of fuzzy thesaurus, *Information Processing & Management*, Vol. 12, pp. 313-318, 1976.
- [11] Wang, Y-C.; Vandendorpe J.; Evans, M. Relational Thesauri in Information Retrieval, *Journal of the American society for information science*, 36(1) pp. 15-27, May, 1985.
- [12] Wu, J. K.; Narasimhalu, A. D; Mehtre, B. M; Lam, C. P; Gao, Y. J. CORE: a content -based retrieval engine for multimedia information systems, *Multimedia Systems*, 3, pp. 25-41, 1995.
- [13] Yager, R. R.; Ovchinnikov, S.; Tong, R. M.; Nguyen, H. T. Fuzzy Sets and Applications: Selected Papers by L. A. Zadeh, John Wiley & Sons, Inc., 1987.
- [14] Carter, S. C.; Kraft, D. H. A generalization and clarification of the waller-kraft wish list, *Information Processing & Management* Vol. 25, No. 1, pp. 15-25, 1989.
- [15] Kraft, D. H.; Bordogna, G.; Pasi, G. An extended fuzzy linguistic approach to generalize boolean information retrieval, *Information Sciences*, 2, pp. 119-134, 1994.
- [16] 최명복, 김민구, 퍼지 언어적 관련도에 근거한 시소러스 모델, 한국정보과학회, 1999년 6월 제출되어 현재 심사 중에 있음. 논문 번호 99-209.
- [17] Kandel, A. Fuzzy mathematical techniques with applications. CA: Addison-Wesley, 1986.
- [18] George J. K.; Tina A. F. Fuzzy Sets, Uncertainty, and Information, Prentice-Hall International, Inc., 1988.
- [19] Chen, S. M.; Ke, J. S.; Chang, J. F. Techniques for handling multicriteria fuzzy decision-making problems, in Proc. 4th International Symposium on Computer and Information Sciences, Cesme, Turkey, vol. 2, pp. 919-925, Oct. 1989.
- [20] Chen, S. M. A weighted fuzzy reasoning algorithm for medical diagnosis, *Decision Support System*, vol. 11, pp. 37-43, 1994.
- [21] Kaufmann, A. and Gupta, M. M. Fuzzy Mathematical Models in Engineering and Management Science. Amsterdam, The Netherlands: Elsevier, 1988.
- [22] Novak, V., Fuzzy Sets and Their Applications, Adam-Hilger, Bristol, 1989.
- [23] Novak, V., A Horizon Shifting Model of Linguistic Hedges for Approximate Reasoning, The Proceedings of the Fifth IEEE International Conference on Fuzzy Systems, pp. 423-427. Volume 1, September 8-11, 1996.



최명복 (Myeong-Bok Choi)

1992년 : 호서대학교 전자계산학과(학사)
 1994년 : 아주대학교 컴퓨터공학과(석사)
 1994년~현재 : 아주대학교 컴퓨터 공학과(박사과정)
 1997년~현재 : 원주대학 행정전산과 (조교수)
 관심분야 : 퍼지응용, 지능형 정보검색, 지식표현, 의사결정 시스템



김민구 (Min-Koo Kim)

1977년 : 서울대학교 계산통계학과 (이학사)
 1979 : 한국과학기술원 전산학과 (공학석사)
 1989년 : Pennsylvania 주립대(박사)
 1981~현재 : 아주대학교 컴퓨터공학과 (교수)

관심분야 : 지식표현, 상식추론, 지능형 교수 시스템, 지능형 정보검색 시스템