

자연어 질의가 가능한 퍼지 기반 지능형 전자상거래 검색 에이전트

Fuzzy Theory based Electronic Commerce Navigation Agent that can Query by Natural Language

김명순* 정환목**
Myeongsoon Kim Hwanmook Chung

* 동주대학 컴퓨터정보통신계열

Dept. of Computer Information Communication, Dongju College

Tel : 051-200-3305 E-mail : mskim@dongju.ac.kr

** 대구가톨릭대학교 컴퓨터정보통신공학부

School of Computer Information Communication, Catholic University of Daegu

Tel : 053-8530-2741 E-mail : hmchung@cuth.cataegu.ac.kr

Abstract

In this paper, we proposed the intelligent navigation agent model for successive electronic commerce management. For allowing intelligence, we used fuzzy theory. Fuzzy theory is very useful method where keywords have vague conditions and system must process that conditions. So, using fuzzy theory, we proposed the model that can process the vague keywords effectively. Through the this, we verified that we can get the more appropriate navigation result than any other crisp retrieval keywords condition.

Key Words : Electronic Commerce System, Fuzzy Membership Value, Navigation Agent, Database, Vague Query, Natural Language Processing

I. 서론

인터넷을 기반으로 하는 전자상거래시스템(electronic commerce system)은 시간과 공간의 제약을 받지 않고 상품 판매가 가능하고 다양한 사용자 인터페이스를 제공함으로써 편리한 구매 및 판매 수단을 제공한다. 이러한 편의성 때문에 전자상거래시스템의 수는 많이 증가했지만 현재 까지도 구매자의 욕구를 100% 만족시키는 시스템의 수는 그렇게 많지 않은 실정이다[1,5,6].

본 연구에서는 전자상거래시스템의 효율적인 운영을 위해 정보 검색 단계에서의 에이전트의 역할에 초점을 맞추어 자연어 질의가 가능한 퍼지 기반 검색 에이전트 모델을 제안한다. 전통적인 마켓(market)에서의 판매원은 구매자들의 "좀 더싼 것은 없나요?"와 같은 요구에 대해 쉽게 검색 서비스를 한다. 그러나, 전자상거래시스템에서는 이와 같은 요구가 불가능하던가 가능하더라도 매우 불만족스러운 결과를 생성한다. 그것은 기존의 검색 에이전트들이 상품 카테고리, 상품명, 제조회사 등의 명사형 검색어(key

word)에 대해서만 제한적으로 검색 기능을 수행했기 때문이다[4]. 따라서, 본 연구에서는 이와는 달리 상품 규격, 가격대 등의 조건에 대하여 「매우크다」, 「크다」, 「보통이다」, 「작다」, 「매우작다」 혹은 「매우비싸다」, 「비싸다」, 「보통이다」, 「싸다」, 「매우싸다」 등과 같은 퍼지(fuzzy) 변수를 고려하고, 이에 적합한 퍼지 귀속도 값(fuzzy membership value)을 적용하여 구매자가 자연어(natural language)를 이용하여 모호한 질의(vague query)를 하더라도 이에 대해 적당한 상품을 제시할 수 있는 시스템 모델을 제안함으로써 불확실한 구매 조건 하에서도 최적의 검색 결과를 제공할 수 있는 방안을 모색한다.

II. 퍼지 전자상거래 검색 에이전트

II-1. 지능형 에이전트

검색 에이전트를 통해 우리는 많은 정보를 얻을 수 있지만 경우에 따라서는 너무나 많은 정보가 제공됨으로써 오히려 구매자들에게 혼란만

가중시키는 결과를 초래할 수도 있다. 따라서, 무조건 많은 정보를 검색할 수 있는 검색 에이전트보다는 구매자의 요구에 정확하게 부응할 수 있는 검색 에이전트의 능력을 더 높히 평가한다. 이를 위해서는 구매자의 검색어를 지능적으로 분석하여 필요한 정보의 범위를 정확하게 추정하는 것이 필요하다. 지능형 에이전트(intelligent agent)란 에이전트가 지능적이라는 말이다. 만일 그 에이전트가 검색 에이전트라면 검색하는 과정이 매우 지능적이라는 것이고, 판매 에이전트라면 판매하는 과정이 지능적이라는 것이다[4]. 결국 지능형 에이전트라고 하는 것은 인간의 지적 능력을 에이전트가 보유하고 있다는 것이고, 지능형 검색 에이전트라면 인간이 자연어로 구사하는 그대로의 질의를 처리할 수 있음을 의미한다. 그러나, 불행하게도 아직 완전한 의미의 자연어 질의 검색 에이전트는 존재하지 않는다. 비슷한 수준에서 자연어 처리를 할 수 있는 방법이 퍼지 이론을 이용하는 것이다[2,3]. 본 연구에서는 퍼지 이론의 한계가 불분명한 변수에 대한 범위를 지정할 수 있다는 것에 착안하여 상품 규격 혹은 가격대 등의 검색 기준에 퍼지적 변수를 사용할 수 있도록 하는 것이다.

II-2. 퍼지 기반 전자상거래 검색 에이전트

II-2-1. 상품 속성 테이블 구성과 정규화

전자상거래 시스템이 운영되기 위해서는 상품과 관련된 데이터들을 데이터베이스 테이블로 구성해야 하는데 다음 (식 1)과 같이 상품 속성 집합이 있다고 가정하자. A_1 은 속성 1을 의미하고 $\{A_{11}, A_{12}, \dots, A_{1m}\}$ 은 속성 1이 가질 수 있는 속성값의 집합이다.

$$A_1 = \{A_{11}, A_{12}, \dots, A_{1m}\}$$

$$A_2 = \{A_{21}, A_{22}, \dots, A_{2m}\}$$

$$\vdots$$

$$A_n = \{A_{n1}, A_{n2}, \dots, A_{nm}\}$$

(식 1)

(단, n은 속성 수, m은 속성값의 수이며 나열순서는 오름차순)
이 때, 속성값 A_{11} 에 대한 정규화 값 $NOR(A_{11})$ 은 다음과 같이 구할 수 있다.

$$NOR(A_{11}) = \frac{A_{11}}{A_{1m}} \times \frac{A_{1m}}{\max} \quad (\text{식 2})$$

(단, max는 존재하는 속성값 중 최대값을 의미)
여기서 max는 상품 속성 테이블에는 없으나 실제로는 존재하는 속성값 중 최대값을 의미하는 것으로 경우에 따라서는 A_{1m} 과 max가 동일할 수 있다. 다른 속성값들도 동일한 방법으로 구할 수 있다.

II-2-2. 상품 속성 값에 대한 퍼지 귀속도

이 중 모호한 질의를 할 수 있는 임의의 속성 k가 있다면 이 속성 A_k 에 대해서 다음과 같은 퍼지 변수 집합 FA_k 를 정의할 수 있다.

$$FA_1 = \{FA_{11}, FA_{12}, \dots, FA_{1m}\}$$

$$FA_2 = \{FA_{21}, FA_{22}, \dots, FA_{2m}\}$$

$$\vdots$$

$$FA_n = \{FA_{n1}, FA_{n2}, \dots, FA_{nm}\}$$

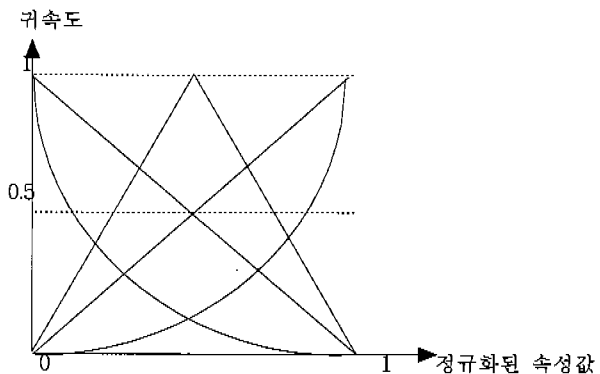
(식 3)

예를 들어, 샴푸의 『단가』라는 속성에 대해서 10000원, 20000원 등 다양한 단가가 있을 수 있지만 이것을 모두 5개의 퍼지 변수로 정의하면 「매우비싸다(VE)」, 「비싸다(E)」, 「보통이다(M)」, 「싸다(C)」, 「매우싸다(VC)」 등으로 표현하는 것이 가능하다. 이 때, 상품의 단가는 모두 5개의 퍼지 변수에 대한 퍼지 귀속도 값을 갖게 된다.

퍼지 변수들에 대한 각 속성 값의 퍼지 귀속도를 결정하기 위해서는 다음 <그림 1>과 같은 5개의 함수를 사용하고 이를 식으로 표현하면 다음과 같다.

$$\begin{aligned} \text{「매우비싸다(VE)」} &= x^2 \\ \text{「비싸다(E)」} &= x \\ \text{「보통이다(M)」} &= \begin{cases} 2x & ; x \leq 0.5 \\ -2(x-1) & ; x > 0.5 \end{cases} \quad (\text{식 4}) \\ \text{「싸다(C)」} &= -x+1 \\ \text{「매우싸다(VC)」} &= (x-1)^2 \end{aligned}$$

(단, x는 속성값들의 정규화 값)



<그림 1> 5개 퍼지 변수에 대한 퍼지 함수

이것은 『규격』을 나타내는 퍼지 변수들 「매우크다(VL)」, 「크다(L)」, 「보통이다(M)」, 「작다(S)」, 「매우작다(VS)」에 대해서도 그대로 적용할 수 있다.

그러면 속성 A_k 에 대한 퍼지 변수 집합 FA_k 는 (식 3)에 의해 $FA_k = \{VE, E, M, C, VC\}$ 가 된다. 또한, 속성 A_k 에 대해서 실제 속성값들은 5개의 퍼지 변수들에 대한 각각의 귀속도를 갖는 5개의 순서쌍 $FMGA_k$ 으로 표현되며 이는 다음 (식 5)와 같다.

$$FMGA_k = \{\mu_{VE}(A_{k1}), \mu_E(A_{k1}), \dots, \mu_{VC}(A_{k1})\} \quad (\text{식 5})$$

(단, $\mu_{VE}(A_{k1})$ 는 퍼지 귀속함수 「VE」에 대한 속성값 A_{k1} 의 퍼지 귀속도를 의미)

만일, 『규격』이라는 속성 A_j 를 고려한다면 「매우크다(VL)」, 「크다(L)」, 「보통이다(M)」, 「작다(S)」, 「매우작다(VS)」라는 5개의 퍼지

변수를 생각할 수 있고, 속성 A_j 에 대한 퍼지 변수 집합 FA_j 는 (식 3)에 의해 $FA_j = \{VL, L, M, S, VS\}$ 가 된다. 또한, 속성 A_j 에 대해서 실제 속성값들은 5개의 퍼지 변수들에 대한 각각의 귀속도를 갖는 5개의 순서쌍 $FMGA_j$ 로 표현되며 이는 다음 (식 6)과 같다.

$$FMGA_j = \{\mu_{VL}(A_{j1}), \mu_L(A_{j1}), \dots, \mu_{VS}(A_{j1})\} \text{ (식 6)}$$

(단, $\mu_{VL}(A_{j1})$ 는 퍼지 귀속함수 'VL'에 대한 속성값 A_{j1} 의 퍼지 귀속도를 의미)

마지막으로, 소비자에 의해 제시된 속성 『규격』, 『단가』 등에 대한 퍼지 변수는 속성값 중 가장 큰 귀속도를 보이는 속성값을 선택하도록 한다. 그러면 상품 속성 『규격』, 『단가』에 대한 퍼지 귀속도가 임계치 H 이상인 결과를 소비자에게 제시함으로써 지능형 에이전트의 임무는 종료된다.

III. 모의 실험 및 고찰

III-1. 실험 상황 설정

본 논문에서 제안하는 모델을 예를 통해 확인해 보자. 다음은 옥실용품을 판매하는 전자상거래시스템이 확보하고 있는 상품 데이터베이스 테이블이다. 단, 속성 『규격』에 대한 max는 1.8, 속성 『단가』에 대한 max는 7000이다.

<표 1> 상품 데이터베이스(T₁)

상품명	규격 (Kg)	제조 회사	제조 국가	단가 (원)	상품명	규격 (Kg)	제조 회사	제조 국가	단가 (원)
삼푸	0.5	한미	한국	2500	삼푸	0.5	국제	한국	2300
삼푸	0.5	한미	한국	2800	삼푸	0.5	국제	한국	2500
삼푸	0.5	한미	한국	3000	삼푸	0.5	국제	한국	3000
삼푸	0.5	한미	한국	3500	삼푸	0.5	국제	한국	3600
삼푸	0.5	한미	한국	4000	삼푸	0.5	국제	한국	4200
삼푸	0.8	한미	한국	2800	삼푸	0.7	국제	한국	2500
삼푸	0.8	한미	한국	3400	삼푸	0.7	국제	한국	3000
삼푸	0.8	한미	한국	4000	삼푸	0.7	국제	한국	3800
삼푸	0.8	한미	한국	4500	삼푸	0.7	국제	한국	5000
삼푸	0.8	한미	한국	5000	삼푸	0.7	국제	한국	5200
삼푸	1	한미	한국	5000	삼푸	1.5	국제	한국	6000
삼푸	1	한미	한국	5500	삼푸	1.5	국제	한국	6200
삼푸	1	한미	한국	6000	삼푸	1.5	국제	한국	6500
삼푸	1	한미	한국	7000	삼푸	1.5	국제	한국	7000

III-2. 퍼지 기반 검색 에이전트의 동작

<표 1>의 테이블에서 퍼지적 표현이 가능한 속성인 『규격』을 A_1 이라고 하고 다음과 같이 5 종류의 퍼지 변수를 적용하자. 퍼지 변수는 『VL』, 『L』, 『M』, 『S』, 『VS』이다. 그러면 속성 A_1 은 다음의 속성값 집합이 된다.

$$A_1 = \{0.5, 0.7, 0.8, 1, 1.5\} \text{ (식 7)}$$

(단, 단위는 Kg)

또한, 속성 A_1 에 대한 퍼지 변수 집합 FA_1 은 다음과 같이 된다.

$$FA_1 = \{VL, L, M, S, VS\} \text{ (식 8)}$$

<표 2>는 다시 이것을 폐구간 [0, 1]로 정규화하여 구한 이들에 대한 퍼지 귀속도이다.

또한, 『단가』 속성을 A_2 라 하면 A_2 에 대한 다음과 같은 5 종류의 퍼지 변수가 만들어지면서 A_2 와 FA_2 가 생성된다. <표 3>은 역시 정규화한 퍼지 귀속도이다.

<표 2> 속성 『규격』에 대한 퍼지 귀속도

규격	0.5	0.7	0.8	1	1.5
정규화값	0.28	0.39	0.44	0.56	0.83
VL	0.08	0.15	0.19	0.31	0.69
L	0.28	0.39	0.44	0.56	0.83
M	0.56	0.78	0.88	0.88	0.34
S	0.72	0.61	0.56	0.44	0.17
VS	0.52	0.37	0.32	0.19	0.03

<표 3> 『단가』에 대한 퍼지 귀속도

단가	정규화값	VE	E	M	C	VC
2300	0.33	0.11	0.33	0.66	0.67	0.45
2500	0.36	0.13	0.36	0.71	0.64	0.41
2800	0.4	0.16	0.4	0.8	0.6	0.36
3000	0.43	0.18	0.43	0.86	0.57	0.33
3400	0.49	0.24	0.49	0.97	0.51	0.26
3500	0.5	0.25	0.5	1	0.5	0.25
3600	0.51	0.26	0.51	0.97	0.49	0.24
3800	0.54	0.29	0.54	0.91	0.46	0.21
4000	0.57	0.33	0.57	0.86	0.43	0.18
4200	0.6	0.36	0.6	0.8	0.4	0.16
4500	0.64	0.41	0.64	0.71	0.36	0.13
5000	0.71	0.51	0.71	0.57	0.29	0.08
5200	0.74	0.55	0.74	0.51	0.26	0.07
5500	0.79	0.62	0.79	0.43	0.21	0.05
6000	0.86	0.73	0.86	0.29	0.14	0.02
6200	0.89	0.78	0.89	0.23	0.11	0.01
6500	0.93	0.86	0.93	0.14	0.07	0.01
7000	1	1	1	0	0	0

$$A_2 = \{2300, 2500, \dots, 6500, 7000\} \text{ (식 9)}$$

(단, 단위는 원)

$$FA_2 = \{VE, E, M, C, VC\} \text{ (식 10)}$$

<표 2>와 <표 3>을 통해 얻은 두가지 속성에 대한 속성값의 퍼지 귀속도를 구하면 다음 <표 4>, <표 5>와 같이 된다. 여기서, 속성 『규격』의 귀속도는 (VL, L, M, S, VS)의 순서이고, 속성 『단가』의 귀속도는 (VE, E, M, C, VC)의 순서이다.

III-3. 질의 테스트 및 결과의 고찰

다음과 같이 구매자의 질의가 있었다고 가정하자.

```
SELECT 상품명 = 삼푸
FROM T1, T2, T3
WHERE 「규격」 = "크다",
「귀속도」 ≥ "0.5" and
「단가」 = "보통이다",
「귀속도」 ≥ "0.25"
```

그러면 지능형 검색 에이전트는 상품 데이터베이스 테이블 T₁로부터 T₂와 T₃을 조인(join)하여 다음과 같은 결과를 도출해준다. 지금과 같은 경우는 질의 조건에 부합되는 즉, 『규격』 속성

이 “크다”이고, 「단가」 속성은 “보통이다”이면 이 두 속성에 대한 퍼지 귀속도가 각각 0.5, 0.25 이상인 상품이 오직 하나 뿐이지만 질의 조건에 따라서는 복수개가 얻어질 수도 있다.

<표 4> 『규격』에 대한 귀속도 테이블(T₂)

상품명	규격	
	규격(Kg)	귀속도
삼푸	0.5	(0.08, 0.28, 0.56, 0.72, 0.52)
삼푸	0.7	(0.15, 0.39, 0.78, 0.61, 0.37)
삼푸	0.8	(0.19, 0.44, 0.88, 0.56, 0.32)
삼푸	1	(0.31, 0.56, 0.88, 0.44, 0.19)
삼푸	1.5	(0.69, 0.83, 0.34, 0.17, 0.03)

<표 5> 『단가』에 대한 귀속도 테이블(T₃)

상품명	단가	
	단가(원)	귀속도
삼푸	2300	(0.11, 0.33, 0.66, 0.67, 0.45)
삼푸	2500	(0.13, 0.36, 0.71, 0.64, 0.41)
삼푸	2800	(0.16, 0.4, 0.8, 0.6, 0.36)
삼푸	3000	(0.18, 0.43, 0.86, 0.57, 0.33)
삼푸	3400	(0.24, 0.49, 0.97, 0.51, 0.26)
삼푸	3500	(0.25, 0.5, 1, 0.5, 0.25)
삼푸	3600	(0.26, 0.51, 0.97, 0.49, 0.24)
삼푸	3800	(0.29, 0.54, 0.91, 0.46, 0.21)
삼푸	4000	(0.33, 0.57, 0.86, 0.43, 0.18)
삼푸	4200	(0.36, 0.6, 0.8, 0.4, 0.16)
삼푸	4500	(0.41, 0.64, 0.71, 0.36, 0.13)
삼푸	5000	(0.51, 0.71, 0.57, 0.29, 0.08)
삼푸	5200	(0.55, 0.74, 0.51, 0.26, 0.07)
삼푸	5500	(0.62, 0.79, 0.43, 0.21, 0.05)
삼푸	6000	(0.73, 0.86, 0.29, 0.14, 0.02)
삼푸	6200	(0.78, 0.89, 0.23, 0.11, 0.01)
삼푸	6500	(0.86, 0.93, 0.14, 0.07, 0.01)
삼푸	7000	(1, 1, 0, 0, 0)

<검색 결과>	
♣ 상품명 : 삼푸	
♣ 규격 : 1.5Kg(귀속도 0.83)	
♣ 단가 : 6,000원(귀속도 0.29)	
♣ 제조회사 : 한미상사	
♣ 제조국가 : 한국	

<그림 2> 검색 결과 예

만일, 전통적인 전자상거래시스템의 검색 에이전트를 이용한다면 앞의 질의 패턴은 불가능하고 다음 <그림 3>과 같은 질의들이 가능하다.

```

SELECT 「상품명」="삼푸"
FROM T1
WHERE 「규격」="1.5Kg" and
「단가」="5,000원~6,000원"

SELECT 「상품명」="삼푸"
FROM T1
WHERE 「규격」="1.5Kg"

SELECT 「상품명」="삼푸"
FROM T1
WHERE 「단가」="5,000원~6,000원"
    
```

<그림 3> 보편적인 검색 예

이 결과로 볼 때, 본 연구에서 제안하는 검색

에이전트 모델은 구매자의 요구사항을 매우 정확하게 파악하여 결과를 도출해 줄 수 있는 반면에 전통적인 검색 에이전트는 적절한 검색 결과를 검색하지 못하든지 아니면 매우 많은 검색 결과를 생성함으로써 구매자로 하여금 판단의 혼란을 초래할 수 있다는 문제점이 있음을 확인할 수 있었다.

IV. 결론

본 연구를 통해 우리는 범위가 명확한 명사형 검색어 환경에서는 검색 결과가 매우 광범위하게 얻어지지만 본 연구에서 제안하고 있는 퍼지 이론을 이용한 지능형 검색 에이전트 모델을 이용하는 경우에는 그 범위가 정확하게 축소됨으로써 구매자에게 판단의 정확성을 높일 수 있는 가치있는 정보를 제공할 수 있음을 확인할 수 있었다. 검색 에이전트의 능력이 검색 결과의 다소로 판단되는 시기는 이미 지났고 이제는 정확하고 유용한 정보를 제공하는 것으로 에이전트의 성능을 결정해야한다. 따라서, 본 연구에서 제안하고있는 모델은 완전한 자연어 검색 에이전트를 개발하기 위한 부품으로 활용할 만한 가치가 충분하다고 사료된다. 그러나, 본 연구에서 제안한 시스템은 일부의 검색 기준에 대해서만 모호한 범위를 처리할 수 있다는 한계가 있었다. 향후로는 구매자가 원하는 모든 검색 기준 항목에 대해서도 모호한 범위를 처리할 수 있도록 하는 연구가 진행되어야 할 것이다.

V. 참고문헌

- [1] 이은석, “에이전트 기술의 전자상거래 응용,” 대한전자공학회지 제 26권 1호, pp.61-70, 1999.
- [2] 최명복, 김민구, “정보검색에서 퍼지 언어 매트릭스에 근거한 효율적인 퍼지 질의 평가 방법,” 한국퍼지및지능시스템학회 논문지 제 10권 3호, pp.218-227, 2000.
- [3] 최명복, 김민구, “정보검색에서 시소러스를 이용한 효율적이고 효과적인 질의 평가 방법,” 한국퍼지및지능시스템학회 논문지 제 10권 6호, pp.605-615, 2000.
- [4] 피수영, 정환목, “지능형 판매에이전트를 이용한 사이버쇼핑물 설계 및 구현,” 한국퍼지및지능시스템학회 논문지 제 10권 5호, pp.497-505, 2000.
- [5] Bhandaru N. and Croft W., “An Architecture for Supporting Goal-based Cooperative Work,” in Gibbs S. and Verriijn-Stuart A., eds., Multi-User Interfaces Applications, pp.337-354, Elsevier Science Publishers B. V., North-Holland, 1990.
- [6] E, S, Lee, “Agent-based Electronic Commerce-Tutorial,” International Conference on Electronic Commerce, April, 1998.