

# 조건부 확률과 퍼지수를 이용한 전자상거래 검색 에이전트 모델

## Electronic Commerce Navigation Agent Model using Conditional Probability and Fuzzy Number

김명순\*      원성현\*\*      정환묵\*\*\*  
Myungsoon Kim, Sunghyun Weon, Hwanmook Chung

\* 동주대학 컴퓨터정보통신계열

Dept. of Computer Information Communication, Donju College

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

\*\* 부산가톨릭대학교 정보공학부

Dept. of Information Engineering, Catholic University of Pusan

Tel : 051-510-0640 E-mail : shwon@cup.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 conditional probability and trapezoidal fuzzy number. Our goal of study is make an intelligent automatic navigation agent model.

**Key Words** : Electronic Commerce System, Fuzzy Number, Trapezoidal Fuzzy Number, Navigation Agent, Profile

### I. 서론

인터넷과 같은 정보 인프라의 확산은 인간의 전통적인 생활 패턴을 큰 폭으로 바꾸는 계기가 되었다. 특히, 상품 구매에 관한 활동은 구매 공간을 가정 혹은 사무실로 옮기게 되었는데 이와 같은 것을 가능하게 하는 소프트웨어를 전자상거래시스템이라하며, 고객들이 보다 편리하게 구매 행위를 할 수 있도록 도와주는 전자상거래시스템의 일부를 에이전트(agent)라고 한다. 그 중 고객을 위해 상품 검색을 도와주는 에이전트를 검색 에이전트라고 한다[1,3,5].

기존의 전자상거래시스템 검색 에이전트들은 고객이 상품 검색을 위해 사용할 수 있는 질의어에 대해 매우 제한적으로 동작해왔다. 즉, 질의 조건은 주로 가격 및 규격 요소에 국한되었고 이 요소들에 대해서도 정확한 질의어(예를 들면, '가격 = 30000원' 혹은 '규격 = 3Kg 이상 4Kg 미만' 등)에 대해서만 검색 기능을 수행할

수 있었다. 이와 같은 검색 에이전트는 고객의 질의 의도를 정확하게 파악하지 못하기 때문에 과다한 검색 결과나 매우 결핍된 검색 결과를 생성하는 단점을 갖고 있다[2,4].

본 논문은 고객이 전자상거래시스템에 접속하여 자신이 원하는 상품을 검색하기 위해 상품명 을 제시했을 때, 해당 고객을 포함한 기존의 고객들의 프로파일 중 고객의 구매 행위에 결정적으로 영향을 미칠 수 있는 요소(여기서는 age, salary, job 등의 3가지 요소)를 선행사건, 구매 성향과 관계된 요소를 후행사건으로 규정하여 고객에 대한 상품 적합도를 계산하고 적합도가 높은 상품 위주로 자동적으로 검색하여 고객에게 제시할 수 있는 검색 에이전트를 설계한다. 이 과정에서 선행사건과 후행사건의 조건부 확률이 계산되어, 후행사건의 선호 구간이 결정되고 선호 구간은 사다리꼴 퍼지수에 의해 퍼지 귀속도가 산출될 것이며 이들을 종합하여 상품

적합도가 계산된다. 고객은 상품 적합도의 임계치(threshold)를 설정함으로써 조건에 맞는 상품 정보를 선택적으로 제공받을 수 있게 된다. 또한, 본 논문에서 제안한 퍼지-확률 속성을 갖는 검색 에이전트 모델은 함수 전문 쇼핑몰의 검색 엔진으로 구현되어 타 쇼핑몰의 검색 에이전트와 비교 분석될 것이다.

## II. 조건부 확률과 퍼지수를 이용한 전자상거래 검색 에이전트 모델

### II-1. 조건부 확률을 이용한 고객 선호 구간

고객의 프로파일에 포함되어 있는 중요 요소를 선행사건으로, 그리고 고객의 성향을 후행사건으로 규정하여 다음 algorithm 1에 의해 최적의 선호구간을 산출한다.

#### algorithm 1

##### step 1

고객의 프로파일 중 중요 신상정보(여기에서는 age, salary, job)를 선행 사건으로 하고 고객의 구매 성향 중 중요 요소(여기에서는 price, size)를 후행 사건으로 규정한다.

##### step 2

각 사건을 분할로 정의한다.

##### step 2-1

선행 사건을 분할(여기에서는 age를 A라 하고 A의 분할은  $A_1, A_2, \dots, A_k$ , salary를 B라 하고 B의 분할은  $B_1, B_2, \dots, B_h$ , job을 C라 하고 C의 분할은  $C_1, C_2, \dots, C_q$ )한다.

##### step 2-2

후행 사건을 분할(여기에서는 price를 X라 하고 X의 분할은  $X_1, X_2, \dots, X_s$ , size를 Y라 하고 Y의 분할은  $Y_1, Y_2, \dots, Y_t$ )한다.

##### step 3

선행 사건(본 논문에서는 A, B, C)에 대해 후행 사건(본 논문에서는 X, Y)이 일어날 확률을 조건부 확률로 구한다. 단,  $X_j(j = 1, 2, \dots, s)$ ,  $Y_j(j = 1, 2, \dots, t)$ ,  $A_i(i = 1, 2, \dots, k)$ ,  $B_i(i = 1, 2, \dots, h)$ ,  $C_i(i = 1, 2, \dots, q)$ 에 대해 각각 구한다.

$$P(X_j | A_i) = \frac{P(A_i \cap X_j)}{P(A_i)}, P(A_i) \neq 0 \quad (\text{식 1})$$

$$P(X_j | B_i) = \frac{P(B_i \cap X_j)}{P(B_i)}, P(B_i) \neq 0 \quad (\text{식 2})$$

$$P(X_j | C_i) = \frac{P(C_i \cap X_j)}{P(C_i)}, P(C_i) \neq 0 \quad (\text{식 3})$$

$$P(Y_j | A_i) = \frac{P(A_i \cap Y_j)}{P(A_i)}, P(A_i) \neq 0 \quad (\text{식 4})$$

$$P(Y_j | B_i) = \frac{P(B_i \cap Y_j)}{P(B_i)}, P(B_i) \neq 0 \quad (\text{식 5})$$

$$P(Y_j | C_i) = \frac{P(C_i \cap Y_j)}{P(C_i)}, P(C_i) \neq 0 \quad (\text{식 6})$$

##### step 4

선행 사건  $A_i(i = 1, 2, \dots, k)$ ,  $B_i(i = 1, 2, \dots, h)$ ,  $C_i(i = 1, 2, \dots, q)$ 에 대한 조건부 확률 값

중 최대값을 각각에 대한 대표값으로 결정한다.

$$SET(A | X) = \bigcup_{i=1}^k (A_i, FACT(MAX(\bigcup_{j=1}^s P(X_j | A_i))))$$

(단,  $SET(A | X)$ 는 A의 분할  $A_1, A_2, \dots$ 에 대하여 최대 조건부 확률을 갖는 항목 X의 집합을 의미하고,  $FACT()$ 는 최대 조건부 확률을 갖는 항목을 의미한다.)

$$SET(B | X) = \bigcup_{i=1}^h (B_i, FACT(MAX(\bigcup_{j=1}^s P(X_j | B_i))))$$

$$SET(C | X) = \bigcup_{i=1}^q (C_i, FACT(MAX(\bigcup_{j=1}^s P(X_j | C_i))))$$

$$SET(A | Y) = \bigcup_{i=1}^k (A_i, FACT(MAX(\bigcup_{j=1}^t P(Y_j | A_i))))$$

$$SET(B | Y) = \bigcup_{i=1}^h (B_i, FACT(MAX(\bigcup_{j=1}^t P(Y_j | B_i))))$$

$$SET(C | Y) = \bigcup_{i=1}^q (C_i, FACT(MAX(\bigcup_{j=1}^t P(Y_j | C_i))))$$

(식 7) ~ (식 12)

##### step 5

step 4로부터 얻은 대표값을 통해 선행 사건들이 선호하는 후행 사건들의 값(구간)을 구한다.

### II-2. 사다리꼴 퍼지수를 이용한 퍼지 귀속도

algorithm 1을 통해 얻은 선호구간에서 고객 성향 요소들에 대한 퍼지 귀속도 값을 얻기 위하여 다음 algorithm 2를 이용한다.

#### algorithm 2

##### step 1

algorithm 1의 step 5로부터 얻은 각 선행사건들의 구간이  $[a_1, a_2]$ ,  $[b_1, b_2]$ ,  $[c_1, c_2]$  등으로 얻어지면 대표 구간은 다음 식으로 구한다.

$$INTERVAL(X) = [\frac{a_1 + b_1 + c_1}{3}, \frac{a_2 + b_2 + c_2}{3}] \quad (\text{식 13})$$

$$INTERVAL(Y) = [\frac{a_1 + b_1 + c_1}{3}, \frac{a_2 + b_2 + c_2}{3}] \quad (\text{식 14})$$

##### step 2

step 1로부터 얻은 구간을 통해 사건 X와 Y에 대한 퍼지 귀속도는 다음과 같은 사다리꼴 퍼지수로 구한다.

$$\mu_X(x) = \begin{cases} 0 & , x < a_1 \\ \frac{x - a_1}{a_2 - a_1} & , a_1 \leq x \leq a_2 \\ 1 & , a_2 \leq x \leq a_3 \\ \frac{a_4 - x}{a_4 - a_3} & , a_3 \leq x \leq a_4 \\ 0 & , x \geq a_4 \end{cases} \quad (\text{식 15})$$

$$\mu_Y(y) = \begin{cases} 0 & , y < a_1 \\ \frac{y - a_1}{a_2 - a_1} & , a_1 \leq y \leq a_2 \\ 1 & , a_2 \leq y \leq a_3 \\ \frac{a_4 - y}{a_4 - a_3} & , a_3 \leq y \leq a_4 \\ 0 & , y \geq a_4 \end{cases} \quad (\text{식 16})$$

##### step 3

step 1에서 결정된 구간 내의 값 중 step 2에 의해 계산된 귀속도를 해당 항목의 적합도라 하고, 이 적합도 중 고객이 제시한 적합도 이상이 되는 상품을 고객에게 자동 검색.

III. 적용 사례

III-1. 데이터베이스 및 데이터 현황

다음은 본 논문에서 제안하는 방법의 모의 테스트를 위해 구축된 데이터베이스 현황이다.

<표 1> 고객 프로파일(신상정보) 데이터베이스

name	ID	sex	age	scholarship	marriage	salary	job

<표 2> 고객의 상품 선호 경향 데이터베이스

name	ID	incense	dcnsity	price	size

<표 3> 상품 데이터베이스(T<sub>1</sub>)

상품명	규격(ml)	농도	제조회사	가격(원)
리브고쉬	50	EDT	이브생로랑	34000
베이비글	50	EDT	이브생로랑	43000
위지	30	EDT	쇼파드	30000
알튀르	50	EDP	샤넬	72000
알튀르	50	EDT	샤넬	66000
알튀르포맨	50	EDT	샤넬	55000
NO.5	50	EDT	샤넬	67000
크로	50	EDT	샤넬	67000
포트올리오(W)	10	EDT	페리엘리스	10000
터치우먼	30	EDP	버버리	34000
터치우먼	50	EDP	버버리	45000
터치포맨	30	EDT	버버리	32000
워크엔드(W)	30	EDT	버버리	27000
워크엔드(W)	50	EDT	버버리	35000
데흐쉬망	30	EDT	나나리찌	30000
미라클	50	EDP	랑콤	62000
뜨레출	30	EDT	랑콤	42000
에센스	50	EDT	베르사체	44000
플라워바이겐조	30	EDT	겐조	36000
플라워바이겐조	50	EDT	겐조	49000
엔비우먼	30	EDT	구찌	31000
엔비우먼	50	EDT	구찌	41000
러쉬포맨	10	EDT	구찌	10000
그랑쌍봉	50	EDT	지방시	29000
그린티오브퍼퓸	75	EDP	엘리자베트아덴	43000
그린티	50	EDT	엘리자베트아덴	32000
5번가	75	EDP	엘리자베트아덴	42000
사파리우먼	75	EDP	랄프로렌	65000
폴로스포르맨	75	EDT	랄프로렌	45000
폴로스포르맨	10	EDT	랄프로렌	12500

III-2. 적합도 산출 및 검색 과정

앞의 algorithm 1, 2에 의해 다음과 같이 적합도가 산출되고 적합한 상품 정보가 검색된다.

algorithm 1

step 1 ~ step 2

- A(age) = {(10대,A<sub>1</sub>), (20대,A<sub>2</sub>), ..., (60대,A<sub>6</sub>)}
- B(salary) = {(1000만원미만,B<sub>1</sub>), (2000만원미만,B<sub>2</sub>), ..., (4000만원미만,B<sub>4</sub>), (4000만원이상,B<sub>5</sub>)}
- C(job) = {(사무직,C<sub>1</sub>), (서비스,C<sub>2</sub>), ..., (농축수산업,C<sub>8</sub>), (학생,C<sub>9</sub>), (무직,C<sub>10</sub>), (기타,C<sub>11</sub>)}
- X(price) = {(20000원미만,X<sub>1</sub>), (35000원미만,X<sub>2</sub>), (50000원미만,X<sub>3</sub>), (75000원미만,X<sub>4</sub>)}
- Y(size) = {(10ml이하,Y<sub>1</sub>), (30ml,Y<sub>2</sub>), (50ml,Y<sub>3</sub>),

(75ml,Y<sub>4</sub>)}

step 3

<사건 A에 대한 X의 조건부 확률 값>

$$\begin{aligned}
 P(X_1 | A_1) &= \frac{4}{11} & P(X_2 | A_1) &= \frac{6}{11} & P(X_3 | A_1) &= \frac{1}{11} \\
 P(X_4 | A_1) &= \frac{0}{11} & P(X_1 | A_2) &= \frac{22}{142} & P(X_2 | A_2) &= \frac{52}{142} \\
 P(X_3 | A_2) &= \frac{46}{142} & P(X_4 | A_2) &= \frac{22}{142} & P(X_1 | A_3) &= \frac{4}{29} \\
 P(X_2 | A_3) &= \frac{8}{29} & P(X_3 | A_3) &= \frac{8}{29} & P(X_4 | A_3) &= \frac{9}{29} \\
 P(X_1 | A_4) &= \frac{1}{17} & P(X_2 | A_4) &= \frac{4}{17} & P(X_3 | A_4) &= \frac{5}{17} \\
 P(X_4 | A_4) &= \frac{7}{17} & P(X_1 | A_5) &= \frac{0}{7} & P(X_2 | A_5) &= \frac{1}{7} \\
 P(X_3 | A_5) &= \frac{5}{7} & P(X_4 | A_5) &= \frac{1}{7} & P(X_1 | A_6) &= \frac{0}{7} \\
 P(X_2 | A_6) &= \frac{2}{7} & P(X_3 | A_6) &= \frac{4}{7} & P(X_4 | A_6) &= \frac{1}{7}
 \end{aligned}$$

사건 B 및 C에 대한 X의 조건부 확률도 마찬가지로 구할 수 있다.

<사건 A에 대한 Y의 조건부 확률 값>

$$\begin{aligned}
 P(Y_1 | A_1) &= \frac{0}{11} & P(Y_2 | A_1) &= \frac{6}{11} & P(Y_3 | A_1) &= \frac{3}{11} \\
 P(Y_4 | A_1) &= \frac{2}{11} & P(Y_1 | A_2) &= \frac{8}{142} & P(Y_2 | A_2) &= \frac{65}{142} \\
 P(Y_3 | A_2) &= \frac{51}{142} & P(Y_4 | A_2) &= \frac{18}{142} & P(Y_1 | A_3) &= \frac{5}{29} \\
 P(Y_2 | A_3) &= \frac{8}{29} & P(Y_3 | A_3) &= \frac{7}{29} & P(Y_4 | A_3) &= \frac{9}{29} \\
 P(Y_1 | A_4) &= \frac{1}{17} & P(Y_2 | A_4) &= \frac{2}{17} & P(Y_3 | A_4) &= \frac{7}{17} \\
 P(X_4 | A_4) &= \frac{7}{17} & P(Y_1 | A_5) &= \frac{0}{7} & P(Y_2 | A_5) &= \frac{3}{7} \\
 P(Y_3 | A_5) &= \frac{4}{7} & P(Y_4 | A_5) &= \frac{0}{7} & P(Y_1 | A_6) &= \frac{0}{7} \\
 P(Y_2 | A_6) &= \frac{1}{7} & P(Y_3 | A_6) &= \frac{5}{7} & P(Y_4 | A_6) &= \frac{1}{7}
 \end{aligned}$$

사건 B 및 C에 대한 Y의 조건부 확률값도 동일한 방법으로 구할 수 있다.

step 4 ~ step 5

- SET(A | X) = {(10대, 35000원미만), (20대, 35000원미만), (30대, 75000원미만), (40대, 50000원이상), (50대, 50000원미만), (60대이상, 50000원미만)}
- SET(B | X) = {(1000만원미만, 20000원미만), (2000만원미만, 35000원미만), (3000만원미만, 50000원미만), (4000만원미만, 75000원미만), (4000만원이상, 35000원미만)}
- SET(C | X) = {(사무직, 35000원미만), (서비스, 35000원미만), (생산직, 20000원미만), (컴퓨터, 35000원미만), (교육직, 75000원미만), (전문직, 75000원미만), (자영업, 35000원미만), (농축수, 20000원미만), (학생, 50000원미만), (무직, 35000원미만), (기타, 50000원미만)}
- SET(A | Y) = {(10대, 30ml), (20대, 30ml), (30대, 75ml), (40대, 50ml), (50대, 50ml), (60대이상, 50ml)}

SET(B | Y) = {(1000만원미만, 30ml), (2000만원미만, 50ml), (3000만원미만, 30ml), (4000만원미만, 75ml), (4000만원이상, 70ml)}

SET(C | Y) = {(사무직, 50ml), (서비스, 50ml), (생산직, 50ml), (컴퓨터, 30ml), (교육직, 75ml), (전문직, 50ml), (자영업, 30ml), (농축수, 10ml이하), (학생, 30ml), (무직, 75ml), (기타, 30ml)}

**algorithm 2**

**step 1**

INTERVAL(X) = {(30대, [50000원, 75000원]), ((3000만원, 4000만원), [50000원, 75000원]), (교육직, [50000원, 75000원])} =  $(\frac{50000원+50000원+50000원}{3}, \frac{75000원+75000원+75000원}{3})$  = [50000원, 75000원]  
 INTERVAL(Y) = {(30대, 75ml), ((3000만원, 4000만원), 75ml), (교육직, 75ml)} =  $\frac{75ml+75ml+75ml}{3} = 75ml$

**step 2 ~ step 3**

상품명	규격		농도	제조회사	가격	
	ml	귀속도			원	귀속도
리브고쉬	50	0.67	EDT	이브생로랑	34000	0.93
베이비들	50	0.67	EDT	이브생로랑	43000	1
위지	30	1	EDT	쇼파드	30000	0.67
알튀르	50	0.67	EDP	샤넬	72000	0.12
알튀르	50	0.67	EDT	샤넬	66000	0.36
알튀르포맨	50	0.67	EDT	샤넬	55000	0.8
NO.5	50	0.67	EDT	샤넬	67000	0.32
코코	50	0.67	EDT	샤넬	67000	0.32
포토폴리오(W)	10	0	EDT	페리엘리스	10000	0
터치우먼	30	1	EDP	버버리	34000	0.93
터치우먼	50	0.67	EDP	버버리	45000	1
터치포맨	30	1	EDT	버버리	32000	0.93
워크엔드(W)	30	1	EDT	버버리	27000	0.47
워크엔드(W)	50	0.67	EDT	버버리	35000	1
데호뤼망	30	0.67	EDT	니나리찌	30000	0.67
미라클	50	0.67	EDP	랑콤	62000	0.52
프레줄	30	1	EDT	랑콤	42000	1
에센스	50	0.67	EDT	베르사체	44000	1
플라워바이젠조	30	1	EDT	젠조	36000	1
글라워바이젠조	50	0.67	EDT	젠조	49000	1
엔비우먼	30	1	EDT	구찌	31000	0.73
엔비우먼	50	0.67	EDT	구찌	41000	1
러쉬포맨	10	0	EDT	구찌	10000	0
그랑생몽	50	0.67	EDT	지방시	29000	0.6
그린티오드퍼플	75	0	EDP	엘리자베트아덴	43000	1
그린티	50	0.67	EDT	엘리자베트아덴	32000	0.8
5년가	75	0	EDP	엘리자베트아덴	42000	1
사파리우먼	75	0	EDP	탈프로렌	65000	0.4
폴로스포르츠맨	75	0	EDT	탈프로렌	45000	1
폴로스포르츠맨	10	0	EDT	탈프로렌	12500	0

**III-3. 검색 예**

age = '30대', salary = '3500만원', job = '교육직'이라는 신상정보를 갖는 고객을 위한 검색 과정을 예로 들어보자. 만일, price에 대한 적합도는 0.4, size에 대한 적합도는 0.4이라 한다면 다음과 같이 검색이 수행된다.

상품명	규격		농도	제조회사	가격	
	ml	귀속도			원	귀속도
알튀르포맨	50	0.67	EDT	샤넬	55000	0.8
미라클	50	0.67	EDP	랑콤	62000	0.52

age = '20대', salary = '1000만원미만', job = '학생'이라는 신상정보를 갖는 고객을 위한 검색 과정을 예로 들어보자. 만일, price에 대한 적합도는 0.7, size에 대한 적합도는 1이라 한다면 다음과 같이 검색이 수행된다.

상품명	규격		농도	제조회사	가격	
	ml	귀속도			원	귀속도
터치우먼	30	1	EDP	버버리	34000	0.93
터치포맨	30	1	EDT	버버리	32000	0.93
엔비우먼	30	1	EDT	구찌	31000	0.73

**IV. 결론**

대부분의 기존 검색 에이전트들이 키워드를 선택하거나 입력하는 수동식 검색 방식을 채택하고 있는 반면에 본 논문에서 제안한 퍼지-확률 기반의 전자상거래시스템 검색 에이전트 모델은 수동 검색 뿐 아니라 자동 검색 기능을 갖는다. 따라서, 기존 고객들의 프로파일을 통해서 분석된 정보를 이용하여 현재 시스템을 사용하고 있는 고객이 무엇을 원하는 것인가를 자동적으로 결정하여 제공한다. 이 때, 퍼지-확률 속성을 이용하여 상품에 대한 고객의 만족 예상 수치를 의미하는 적합도를 계산하여 적합도가 높은 상품 위주로 검색하도록 한다. 따라서 대부분의 검색 에이전트들이 갖는 단점 중의 하나인 상품 검색시 과도한 검색 결과 혹은 매우 결핍된 검색 결과를 보이지 않고 최적의 검색 결과만을 제공할 수 있는 장점이 있다.

향후에는 사다리꼴 퍼지수의 파라미터가 본 논문에서와 같이 고정되어 있는 것이 아니라 데이터의 변화에 적응적으로 변화할 수 있는 방법을 모색해야 할 것이다. 이 문제는 유전자 알고리즘의 특성을 이용하면 해결 가능할 것으로 사료된다.

**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] 피수영, 정환목, "지능형 판매에이전트를 이용한 사이버쇼핑몰 설계 및 구현," 한국퍼지및지능시스템학회 논문지 제 10권 5호, pp.497-505, 2000.
- [4] Bhandaru N. and Croft W., "An Architecture for Supporting Goal-based

Cooperative Work," in Gibbs S. and Verrijn-Stuart A., eds., Multi-User Interfaces Applications, pp.337-354, Elsevier Science Publishers B. V., North-Holland, 1990.

[5] E, S, Lee, "Agent-based Electronic Commerce-Tutorial," International Conference on Electronic Commerce, April, 1998.