

퍼지-조건부확률을 이용한 전자상거래 검색 에이전트 모델에 관한 연구

김 명 순*

A Study on Electronic Commerce Navigation Agent Model Using Fuzzy-Conditional Probability

Myeongsoon Kim *

요 약

기존의 전자상거래시스템 검색 에이전트들은 고객이 상품 검색을 위해 사용할 수 있는 질의어에 대해 매우 제한적으로 동작해왔다.

본 논문은 고객이 전자상거래시스템에 접속하여 자신이 원하는 상품을 검색하기 위해 상품명을 제시했을 때, 해당 고객을 포함한 기존의 고객들의 프로파일 중 고객의 구매 행위에 결정적으로 영향을 미칠 수 있는 요소를 선행사건, 구매 성향과 관계된 요소를 후행사건으로 규정하여 고객에 대한 상품 적합도를 계산하고 적합도가 높은 상품 위주로 자동적으로 검색하여 고객에게 제시할 수 있는 퍼지-조건부 확률을 이용한 전자상거래 검색 에이전트를 제시한다.

Abstract

In this paper, we proposed the intelligent navigation agent model for successive electronic commerce management. For allowing intelligence, we used fuzzy conditional probability and trapezoidal. 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. Our goal of study is make an intelligent automatic navigation agent model.

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

1. 서론

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 질의어에 대해서만 검색 기능을 수행할 수 있었다. 이와 같은 검색 에이전트는 고객의 질의 의도를 정확하게 파악하지 못하기 때문에 과도한 검색 결과나 매우 결핍된 검색 결과를 초래하는 단점을 갖고 있다 [1][2][3].

본 논문에서는 고객이 전자상거래시스템에 접속하여 자신이 원하는 상품을 검색하기 위해 상품명을 제시했을 때, 해당 고객을 포함한 기존의 고객들의 프로파일(profile) 중 고객의 구매 행위에 결정적인 영향을 미칠 수 있는 요소를 선행사건(prior event), 구매 성향과 관계된 요소를 후행사건(posterior event)으로 규정하여 고객에 대한 상품 적합도를 언어 변수로 제시하고 적합도가 높은 상품 위주로 자동적으로 검색하여 고객에게 제공할 수 있는 검색 에이전트를 설계한다 [4][5]

II. 상품 자동 검색 에이전트 모델의 설계

1. 상품 자동 검색 에이전트 모델의 구성

본 논문에서 제안된 퍼지-조건부확률 기반 상품 자동 검색 에이전트 모델은 (그림 1)과 같이 구성되어 있다.

2. 조건부 확률을 이용한 적합도 산출 알고리즘

상품 속성 선호 구간 결정 알고리즘은 다음과 같은 단계를 거친다.

2-1 Step 1

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

2-2 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.

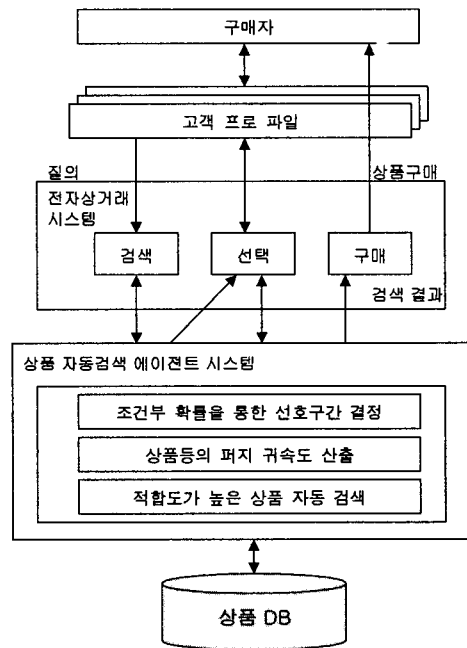


그림 1. 제안된 퍼지-조건부확률 상품 자동 검색 에이전트 모델
Fig 1. Product automatic navigation agent model of proposed fuzzy-conditional probability

• 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.

2-3 Step 3

선행 사건(본 논문에서는 A, B, C)에 대해 후행 사건(본 논문에서는 X, Y)이 일어날 확률을 조건부 확률로 구

한다. 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.

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

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

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

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

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

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

2-4 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 A 의 분할 A_1, A_2, \dots 에 대하여 최대 조건부 확률을 갖는 항목 X 의 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)

2-5 Step 5

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

III. 적용 사례

1. 데이터베이스 및 데이터 현황

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

표 1. 고객 프로파일(산상정보) 데이터베이스
Table1. Database of customer's profile

name	ID	sex	age	scholarship	marriage	salary	job

표 2. 고객의 상품 선호 경향 데이터베이스
Table2. Database of purchasing tastes

name	ID	incense	density	price	size

표 3. 상품데이터베이스(T1)
Table 3. Product Database

상품명	규격(ml)	농도	제조사	가격(원)
리브고쉬	50	EDT	이브생로랑	34000
베이비들	50	EDT	이브생로랑	43000
위지	30	EDT	쇼피드	30000
알뤼르	50	EDP	샤넬	72000
알뤼르	50	EDT	샤넬	66000
알뤼르포맨	50	EDT	샤넬	55000
NO.5	50	EDT	샤넬	67000
코코	50	EDT	샤넬	67000
포도폴리오(W)	10	EDT	페리엘리스	10000
터치우먼	30	EDP	버버리	34000
터치우먼	50	EDP	버버리	45000
터치포맨	30	EDT	버버리	32000
위크엔드(W)	30	EDT	버버리	27000
위크엔드(W)	50	EDT	버버리	35000
데흐뒤땅	30	EDT	나나리찌	30000
미라클	50	EDP	랑콤	62000

드레졸	30	EDT	람콤	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
그린티오드퍼품	75	EDP	엘리자베드아덴	43000
그린티	50	EDT	엘리자베드아덴	32000
5번가	75	EDP	엘리자베드아덴	42000
사파리우먼	75	EDP	랄프로렌	65000
폴로스포르츠맨	75	EDT	랄프로렌	45000
폴로스포르츠맨	10	EDT	랄프로렌	12500

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

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

2. 적합도 산출 및 검색

제 2장의 알고리즘 1에 의해 적합도 산출 및 적합도 결과물 검색하는 과정과 상품 속성 선호 구간 결정은 다음과 같다.

2-1 상품 속성 선호 구간의 결정

• Step 1 ~ Step 2

- A(age) = {(10대, A₁), (20대, A₂), ..., (60대, A₆)}
- B(salary) = {(1000만원미만, B₁), (2000만원미만, B₂), ..., (4000만원미만, B₄), (4000만원이상, B₅)}
- C(job) = {(사무직, C₁), (서비스, C₂), ..., (농축수산업, C₈), (학생, C₉), (무직, C₁₀), (기타, C₁₁)}
- X(price) = {(20000원미만, X₁), (35000원미만, X₂), (50000원미만, X₃), (75000원미만, X₄)}
- Y(size) = {(10ml이하, Y₁), (30ml, Y₂), (50ml, Y₃), (75ml, Y₄)}

• Step 3

(사건 A에 대한 X의 조건부 확률 값)

$$\begin{aligned}
 P(X_1 | A_1) &= \frac{4}{11} & P(X_2 | A_1) &= \frac{6}{11} \\
 P(X_3 | A_1) &= \frac{1}{11} & P(X_4 | A_1) &= \frac{0}{11} \\
 P(X_1 | A_2) &= \frac{22}{142} & P(X_2 | A_2) &= \frac{52}{142} \\
 P(X_3 | A_2) &= \frac{46}{142} & P(X_4 | A_2) &= \frac{22}{142}
 \end{aligned}$$

(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) = {(A1, X2), (A2, X2), (A3, X4), (A4, X4), (A5, X3), (A6, X3)}
- SET(A | X) = {(10대, 35000원미만), (20대, 35000원미만), (30대, 75000원미만), (40대, 75000원미만), (50대, 50000원미만), (60대이상, 50000원미만)}
- SET(B | X) = {(B1, X1), (B2, X2), (B3, X3), (B4, X4), (B5, X2)}
- SET(B | X) = {(1000만원미만, 20000원미만), (2000만원미만, 35000원미만), (3000만원미만, 50000원미만), (4000만원미만, 75000원미만), (4000만원이상, 35000원미만)}
- SET(C | X) = {(C1, X2), (C2, X2), (C3, X1), (C4, X2), (C5, X4), (C6, X4), (C7, X2), (C8, X1), (C9, X3), (C10, X2), (C11, X3)}
- 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)}

2-2 상품에 대한 적합도 산출

• Step 1

- INTERVAL(X) = MAX({(30대, [50000원, 75000원], 7/17), ((3000만원, 4000만원), [50000원, 75000원], 9/19), (교육직, [50000원, 75000원], 12/31)}) = {50000원, 75000원}
- INTERVAL(Y) = MAX({(30대, 75ml, 9/29), ((3000만원, 4000만원), 75ml, 8/19), (교육직, 75ml, 14/31)}) = 75ml

• Step 2

〈표 4〉 선호 가격(50000원~75000원)내의 상품 검색
Table 4. Product navigation in preference price

상품명	규격(ml)	농도	제조회사	가격(원)
알류르	50	EDP	샤넬	72000
알류르	50	EDT	샤넬	66000
아루르포맨	50	EDT	샤넬	55000
NO.5	50	EDT	샤넬	67000
코코	50	EDT	샤넬	67000
미라클	50	EDP	랑콤	62000
사파리우먼	75	EDP	랄프로렌	65000

표 5. 선호 규격(75ml)내의 상품 검색

Table 5. Product navigation in preference size

상품명	규격(ml)	농도	제조회사	가격(원)
그린티오드	75	EDP	지방시	43000
5번가	75	EDP	엘리자베트아덴	42000
사파리우먼	75	EDP	랄프로렌	65000
플로스포츠맨	75	EDT	랄프로렌	45000

• Step 3

단계 1, 2의 결과를 통해 결정된 선호 구간 내에 포함되는 상품의 속성에 대한 귀속도를 다음과 같은 사다리꼴 퍼지수로 계산한다.

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

예) 가격이 72000원인 상품의 귀속도의 계산은 다음과 같다.

(a1, 20000미만), (a2, 35000미만), (a3, 50000미만), (a4, 75000미만)이므로

$$a_3 \leq x \leq a_4, \frac{75000 - 72000}{75000 - 50000} = 0.12 \text{ 이다.}$$

3. 검색 예

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

표 6. 적합도 변수(가격)를 고려한 상품 검색

Table 6. Product navigation considering degree of satisfaction variable(price)

상품명	규격		농도	제조회사	가격	
	ml	귀속도			원	적합도 변수
아루르포맨	50	0.67	EDT	샤넬	55000	비싸다
미라클	50	0.67	EDP	랑콤	62000	비싸다
사파리우먼	75	0	EDP	랄프로렌	65000	비싸다

표 7. 적합도 변수(가격, 규격)를 고려한 상품 검색

Table 7. Product navigation considering degree of satisfaction variables(price, size)

상품명	규격		농도	제조회사	가격	
	ml	적합도 변수			원	귀속도
아루르포맨	50	크다	EDT	샤넬	55000	비싸다
미라클	50	크다	EDP	랑콤	62000	비싸다

IV. 결론 및 향후 연구

기존의 전자상거래시스템 검색 에이전트들은 질의 조건을 주로 가격 및 규격 요소에 국한되었고 이 요소들에 대해서도 정확한 질의어에 대해서만 검색 기능을 수행할 수 있었다. 즉, 고객의 질의 의도를 정확하게 파악하지 못하기 때문에 과도한 검색 결과나 매우 결핍된 검색 결과를 초래하는 단점을 갖고 있고, 대부분의 현존 검색 에이전트들이 키워드를 선택하거나 입력하는 수동식 검색 방식을 채택하고 있기 때문에 고객들이 사용하는데 상당한 불편을 초래하는 결과를 보여주고 있다.

본 논문에서 제안한 상품 자동 검색 에이전트 모델은 고객이 전자상거래시스템에 접속하여 자신이 원하는 상품을 검색하기 위해 상품명을 제시했을 때, 해당 고객을 포함한 기존의 고객들의 프로파일(profile) 중 고객의 구매 행위에 결정적인 영향을 미칠 수 있는 요소를 선행 사건(prior event), 구매 성향과 관계된 요소를 후행 사건(posterior event)으로 규정하여 고객에 대한 상품 적합도를 언어 변수로 제시하고 적합도가 높은 상품 위주로 자동적으로 검색하여 고객에게 제공할 수 있는 검색 에이전트를 설계한다. 이 과정에서 선행사건과 후행 사건의 조건부 확률이 계산되어, 후행 사건의 선호 구간이 결정되고, 사다리꼴 퍼지수(trapezoidal fuzzy number)에 의해 퍼지 귀속도(fuzzy membership value)가 산출이 되며, 이들을 종합하여 상품 적합도가 계산된다. 고객은 상품 적합도의 임계치(threshold)를 언어 변수로 설정함으로써 조건에 맞는 상품 정보를 선택적으로 제공받을 수 있게 된다.

따라서 대부분의 검색 에이전트들이 갖는 단점 중의 하나인 상품 검색 시 과도한 검색 결과 혹은 매우 결핍된 검색 결과를 보이지 않고 최적의 검색 결과만을 제공할 수 있는 장점을 가지고 있다.

향후 연구로는 검색 결과에 대한 고객의 만족도를 조사하여 고객이 상품 검색 시 최적의 검색 결과로 만족할 수 있는 검색 에이전트를 만드는 것이 필요 할 것이다.

참고문헌

- [1] 백혜정 외 5인, "적용형 에이전트," 한국정보과학회지 제 15권 3호, pp.29-38, 1997.
- [2] 신봉기, "인터넷 에이전트," 대한전자공학회지 제 26권 1호, pp.15-24, 1999.
- [3] 이은석, "에이전트 기술의 전자상거래 응용," 대한전자공학회지 제 26권 1호, pp.61-70, 1999.
- [4] 피수영, 정환목, "지능형 판매에이전트를 이용한 사이버 쇼핑물 설계 및 구현," 한국퍼지및지능시스템학회 논문지 제 10권 5호, pp.497-505, 2000.
- [5] E. B. Doorenbos, O. Etzioni and D. S. Weld, "A Scalable Comparison-Shopping Agent for the WWW," 1st International Conference on Autonomous Agent, 1997.
- [6] <http://bf.cstar.ac.com/bf>
- [7] <http://www.ece.curtin.edu.au/~saounb>
- [8] <http://www.dimos.de/dnns/original>

저자소개

김명순



1985년 방송대학교 경영학과 졸업
 1988년 경성대학교 대학원 산업정보학과(공학석사)
 2001년 대구 가톨릭대학교 대학원 전산통계학과(이학박사)
 1993년~현재 동주대학 컴퓨터 정보 제어 조교수
 <관심 분야> 인공지능, 퍼지 이론, 신경망, 다치논리, 자연어 처리