

온톨로지 기반의 지능형 추천 에이전트

조범수^o 김재원 노상욱
가톨릭대학교 컴퓨터 정보 공학부
{cbs1130^o, voards, sunoh}@catholic.ac.kr

Intelligent Recommendation Agent Based on Ontology

Beonsoo Cho^o Jaewon Kim Sanguk Noh
School of Computer Science and Information Engineering
The Catholic University of Korea

요 약

최근 들어 인터넷의 급속한 발전으로 사용자가 처리해야 할 정보의 양이 급속히 늘어나게 됨으로써 사람이 혼자만의 힘으로 이 많은 정보를 처리하는 것이 하나의 고단한 작업이 되었고, 이 작업을 돕기 위한 소프트웨어 에이전트 (software agent)의 필요성이 대두되었다. 본 논문에서는 구현한 소프트웨어 에이전트가 사용자의 업무보조 (personal assistant)라는 자신의 임무를 수행하기 위하여 온톨로지 (ontology)를 기반으로 사용자의 선호도 (preference)와 의사결정 패턴을 학습하여 사용자 프로파일 (user profile)을 작성한다. 학습한 프로파일을 바탕으로 사용자의 선호도와 일치하는 제품을 추천하는 지능형 에이전트를 제안하고, 실질적인 실험을 통해 학습된 사용자의 성향을 분석한다.

1. 서 론

90년대 들어서 인터넷이 보편화 되면서 사람들의 생활에서 온라인이 차지하는 비중이 매우 커지게 되었다. 사람들은 온라인상에서 의사소통을 하고, 쇼핑을 하며 여러 가지 경제활동을 한다. 인터넷이 발달하면서 사용자가 처리해야 할 정보의 양이 많아짐에 따라 이러한 사용자의 수고를 효과적으로 보조해 주는 개인적인도우미 (personal assistant)로서의 자율적인 소프트웨어 에이전트 (autonomous software agent)의 필요성이 점차 증대되고 있다 [1]. 이러한 각각의 에이전트들이 저마다 자신들만의 사용자를 위해 다른 에이전트와 정보를 공유하며 자동적으로 동작하기 위해서는, 정보의 규격화와 사용자의 특성을 학습하는 기능이 반드시 필요하다. 본 논문에서는 이러한 진보적인 에이전트를 구현하기 위하여, 여러 사용자들이 존재하는 인터넷 환경에서 상품에 대한 사용자의 선호특성을 학습하여, 사용자가 원하는 상품을 추천하여 주는 지능형 에이전트를 구현할 것이다. 정보의 규격화를 위하여 온톨로지 (ontology)를 사용하며, 온톨로지 기반의 지능형 추천 에이전트 (intelligent recommendation agent)가 사용자의 선호도를 얼마나 잘 반영하는가를 실험을 통하여 평가할 것이다.

2. 온톨로지 생성기

본 논문에서 온톨로지는 여러 개의 에이전트가 존재하는 환경에서 에이전트간의 의사소통을 위하여, '규격화된 정보의 형태를 표현하는 방식'으로 정의한다 [2]. 본 논문에서는 하나의 예제 도메인으로 디지털 카메라 (digital camera)를 구매하는 인터넷 환경에서 각각의 구매자와 공급자, 또는 마켓의 관리자 사이의 정보공유를 위하여 온톨로지를 생성하였다. 이러한 온톨로지는 디지털 카메라의 특성과 제원을 계층적인 트리 구조 (hierarchy of concepts)로 표현하는데, 이는 상품 공

급자가 하나의 카메라에 대한 인스턴스 (instance)를 생성하는데 기본 틀로 이용되며, 또한 구매자의 선호특성을 학습하여 프로파일을 작성하는데 사용된다. 온톨로지 생성을 위한 도구로는 기존의 Protege-2000, Ontolingua-1997, Chimaera-2000 [2]이 존재하지만, 본 논문에서는 속성 (attribute)의 정의가 자유로움과 속성들 간의 계층구조를 쉽게 표현하는 온톨로지 생성기 (ontology generator)를 구현하였다. 디지털 카메라 도메인에 대한 온톨로지는 카메라의 재원표를 기초로 하여, 각 항목으로부터 비슷한 개념의 항목끼리 그룹화하여 보다 일반적인 개념으로 묶어 나가는 상향식 지식 처리과정에 의해 생성되었다 [2].

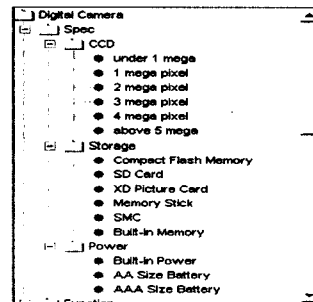


그림 1. 온톨로지 생성기에 의해 구축된 디지털 카메라의 계층적 지식표현.

구현한 온톨로지 생성기를 사용하여 생성한 디지털 카메라의 계층적 지식표현은 그림 1과 같다. 그림 1에서, 최상위 레벨의 '디지털 카메라'는 루트 클래스 (root class)를 나타내고, 두 번째 레벨의 각각의 항목 - 예를 들면, Spec 등 - 들은 세 번째 레벨의 속성 (attribute)들을 그룹으로 묶어주는 추상적인 개념을 나타내며, 최하위 레벨의 노드들은 부모노드인 속성들이 가질 수 있

는 값들을 슬롯 (slot)의 형태로 나타낸다. 가격이나 무게와 같은 숫자를 슬롯으로 가지는 경우에는 지식 노드를 하나만 가지게 되며, 최대치를 초가값으로 지정해야 한다. 이러한 온톨로지의 작성은 마켓 관리 에이전트 (market provider) 에 의해 수행되어 진다.

3. 사용자의 선호도를 학습하는 방법

추천 에이전트 (recommendation agent)의 임무 중에서 사용자의 선호특성을 파악하는 것은 매우 중요한 작업이다. 다속성 유틸리티 이론 (Multi-Attribute Utility Theory: MAUT) 은 소비자가 상품을 평가하는데 있어서 널리 사용되는 이론으로, 본 논문에서는 디지털 카메라 도메인에서 사용자 관점에 따른 상품의 가치(value)를 평가하기 위해 사용하였다 [3].

$$V(x|t) = \sum_{i=1}^n W_i(x|t) \quad (1)$$

식 (1) 은 사용자의 선호도에 따른 상품가치(Value)의 평가함수이다. 특정한 소비자가 주어진 시각 t 에서 고려하는 상품들 중의 하나가 x 이며, $W_i(x|t)$ 는 시각 t 에서 상품 x의 i 번째 속성에 대한 가중치를 나타낸다.

식 (1)은 일반적으로 $\sum_{i=1}^n W_i = 1$ 과 함께 사용되어 상품에 대한 속성별 가중치의 총 합이 1이 되도록 설정이 되지만, 본 논문에서는 가중치 증가를 통한 학습을 위해 이를 제외 하였다.

사용자의 선호도를 학습하기 위하여, 사용자로 하여금 추천된 카메라에 대하여 '좋음' (good) 과 '나쁨' (no good)으로 평가하게 한다 [4]. 이때, '좋음' 으로 선택된 카메라의 각 속성들을 이용하여, 식 (2)에 따라 사용자 프로파일의 가중치 값을 변화시킨다.

$$W_i(x|t) = W_i(x|t-1) + \Delta W \quad (2)$$

이때, ΔW 는 '좋음' 일때 +0.5, '나쁨' 일때 +0 의 변화량을 주었다.

실제 예로, 디지털 카메라 도메인에서 <표 1> 과 같은 상품이 있다고 가정하자.

<표 1> 디지털 카메라 도메인에서 상품 예시

	CCD 화소	광학 줌	제조사	평가
camera a	2백만화소	3×	sony	No Good
camera b	3백만화소	5×	samsung	Good
camera c	2백만화소	None	samsung	No Good
camera d	3백만화소	5×	canon	No Good

<표 1>에서와 같이 네 개의 상품이 준비되어 있을 때 처음 사용자의 성향은 모두 0이라고 가정하고, 위와 같은 평가를 하였다면, '좋음' 으로 평가받은 카메라 b가 갖고 있는 각각의 슬롯에 대한 사용자 프로파일 (user profile) 의 가중치가 0.5씩 증가하여 다음 수행에서는 각 카메라에 대한 V 값은 식 (1) 에 의해 다음과 같이 구할 수 있다.

$$V(a|t) = 0 + 0 + 0 = 0$$

$$V(b|t) = 0.5 + 0.5 + 0.5 = 1.5$$

$$V(c|t) = 0 + 0 + 0.5 = 0.5$$

$$V(d|t) = 0.5 + 0.5 + 0 = 1.0$$

결과적으로, 네 개의 상품에 대한 전체 유틸리티 값 중에서 카메라 b 에 대한 전체 가치값($V(b|t)=1.5$) 이 가장 커지게 된다.

4. 시스템 구조

본 연구를 위해 개발한 지능형 추천 에이전트는 여러 개의 에이전트가 존재하는 환경에서 사용자 에이전트와 구매자 에이전트의 온톨로지를 구축하고, 이를 기반으로 사용자의 선호도를 추적하여 사용자의 특성에 맞는 제품을 추천하는 응용 프로그램 [5] 으로 분류할 수 있다. 시스템의 전체적인 구성이 그림2 에 나와 있다.

마켓관리 에이전트 (market provider) 는 특정 상품에 대한 온톨로지를 온톨로지 생성기를 통해서 만들어서 저장하고 있게 된다. 생성되어진 온톨로지는 커뮤니케이션 모듈 (Communication Module)을 통해서 추천 에이전트나 공급 에이전트에 전송하여 준다. 마켓 관리 에이전트가 서버의 역할도 함께 하고 있기 때문에, 추천 에이전트나 공급 에이전트는 이 에이전트를 통해 통신을 하게 된다.

공급 에이전트 (merchant agent) 는 마켓 관리 에이전트로부터 받은 온톨로지를 기반으로 해서 상품 리스트를 생성하고, 저장한다. 상품리스트는 사용자의 요청이 있을 때, 추천 에이전트에게 전달해 준다.

추천 에이전트 (recommendation agent) 는 사용자의 선호도를 학습하기 위한 학습 모듈 (learning module), 공급 에이전트들로부터 전송받은 상품들을 사용자의 선호도를 이용해서 사용자에게 추천해주는 필터링 모듈 (filtering module), 다른 에이전트들과의 의사소통을 위한 커뮤니케이션 모듈로 구성되어 있다.

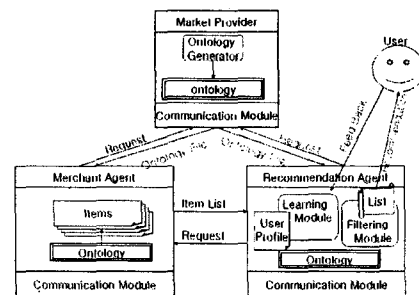


그림 2 마켓관리 에이전트, 공급 에이전트, 지능형 추천 에이전트로 구성된 시스템.

5. 실험

본 논문에서 수행한 실험의 목적은 지능형 추천 에이전트가 축적한 사용자의 선호특성을 바탕으로 새로운 상품이 나왔을 경우, 얼마나 사용자의 선호도에 맞는 상품들을 추천하여 줄 수 있는가를 평가하는 것이다.

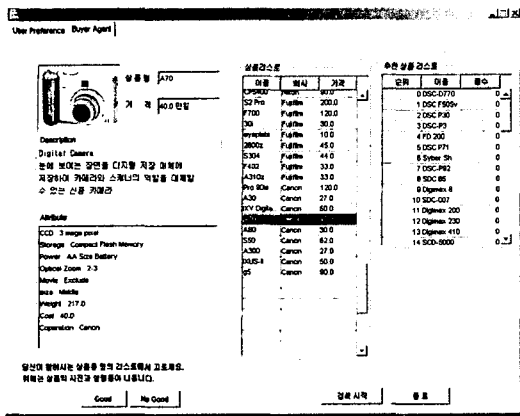


그림 3. 구현한 추천 시스템의 사용자 인터페이스.

그림 3과 같은 화면의 인터페이스를 통해 준비된 100개의 카메라를 대상으로 취향에 맞으면 '좋음', 맞지 않으면 '나쁨'을 선택함으로써 선택된 상품의 해당하는 항목의 가중치를 올려주었다.

학습을 통해 얻어진 사용자 선호도의 가중치 값은 베이즈 확률 분류기(naive Bayesian classifier) [6]로 분석한 결과와의 비교를 위해 0과 1의 범위로 수정(현재 가중치 값 / 데이터에 포함된 해당 속성의 총합)하였고, 그림 4의 그래프에 그 결과를 나타내었다. 베이즈 확률 분류기의 그래프는 사용자가 '좋음'의 평가를 내린 카메라에 대해 속성별로 그룹화한 확률 값을 나타낸다.

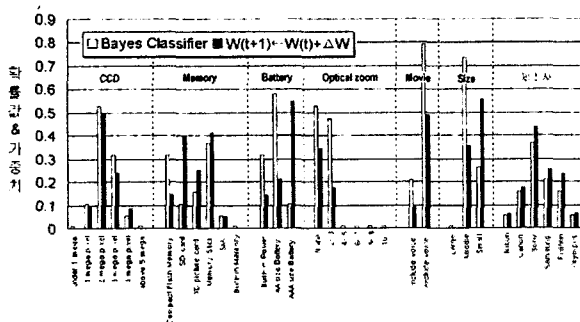


그림 4. 계산한 가중치와 naive Bayesian Classifier를 통하여 얻은 각 속성의 중요도에 대한 결과 비교.

그림 4는 10여명의 사용자 중에서 선택된 한명의 실험 참가자의 선택 결과를 보여준다. 선택된 참가자는 100개의 카메라중 19개의 카메라에 '좋음'의 평가를 했다. 그래프를 요약하면 (1) CCD항목에서는 베이즈 확률과 가중치 증가 값 모두 비슷한 분포로 2백만 화소대의 제품을 가장 선호했다는 것을 알 수 있으며, (2) 3배 이하의 저배율 줌을 채택하고, (3) 크기가 비교적 작은, (4) 'Sony 카메라'를 선호했다. 결과적으로 이 실험자는 가격이 저렴한 보급형의 카메라에 중점을 두고 선택하였다는 것을 알 수 있다.

대부분의 실험 참가자들의 결과 값을 비교하여 정리

한 결과, 선택한 결과 값들이 베이즈 확률 분류기와 가중치 증가를 통한 학습결과가 비슷한 성향을 보여주었으나, 후보 상품들의 실제 값이 특정한 속성에 집중된 경우에는 가중치를 계산한 결과 값과 베이즈 확률의 결과 값이 다른 성향을 보여 주었다. 예를 들어, 그림 4에서 'CCD' 항목과 '제조사' 항목들은 속성의 분포가 균등하여 두 가지 결과값이 비슷하게 나타났다. 반면에, Battery 항목과 Size 항목에서는 준비된 카메라 상품들이 특정한 속성 - 즉, AA 배터리, Medium 크기 - 에 집중되어 있기 때문에, 집중되어 있지 않은 속성에 대해서 '좋음'의 평가를 한번만 내리더라도 가중치 증가 값이 1에 근접하게 된다. 크기 항목에서 중간크기의 카메라를 보면 준비된 데이터에서의 중간크기 카메라의 비율이 높기 때문에, 확률 값과 비교해 가중치의 값이 상대적으로 낮은 수치로 나타나는 것이다.

6. 결론 및 향후 과제

본 논문에서 구현한 추천 시스템에 대한 사용자 만족도는 실험결과 높게 나타났으며, 사용자의 선호도와 일치하는 상품을 추천해 주는 것으로 나타났다. 이는 사용자가 수많은 상품 중에서 자신에게 가장 알맞은 상품을 실질적으로 검색 할 수 있게 해 줌으로써, 사용자로 하여금 어떤 상품을 구매할 것인가에 대한 도우미 역할을 한다. 또한, 온톨로지를 통해 시장 환경에 대한 정보를 공유함으로써, 사용자가 모르는 제품에 대해서도 해당 제품의 사양을 비교할 수 있다는 것 또한 우리 시스템의 장점이라 할 수 있다.

향후 과제로는 현재 다속성 유틸리티 이론과 가중치 증가를 통한 기초적인 형태의 학습을 적용시켰지만, 이를 확장하여 사용자의 일반적인 성향을 파악할 수 있도록 다양한 도메인에 적용시킬 것이다. 또한, 사용자의 피드백이 추천 에이전트에게만 전달되는 것이 아니라 공급 에이전트에게도 전달되도록 함으로써, 공급자가 사용자들의 성향을 한눈에 파악 할 수 있도록 하는 연구를 수행할 것이다.

7. 참고문헌

- [1] M.N. Huhns and M.P. Singh, Personal Assistants, IEEE Internet Computing, pp. 90-92, 1998.
- [2] N.F. Noy and D.L. McGuinness, Ontology Development 101:A guide to Creating Your First Ontology, Stanford, TR KSL-01-05, 2001.
- [3] R. Schaefer, Rules for Using Multi-Attribute Utility Theory for Estimating a User's Interests. In Proc. of the 9th GI-Workshop, 2001.
- [4] F. Menczer, W.N. Street, N. Vishwakarma, A.E. Monge, M. Jakobsson, IntelliShopper:A Proactive, Personal, Private Shopping Assistant, ACM, 2002.
- [5] M.J. Owen, L. Lee, G. Sewell, S. Steward, and D. Thomas, Multi-agent trading environment, BT Technology Journal, Volume 17, Issue 3, 1999.
- [6] S. Noh and P.J. Gmytrasiewicz, Towards Flexible Multi-Agent Decision-Making Under Time Pressure, In Proceedings of IJCAI, pp. 492-498, 1999.