

기계학습 기반 적응형 전자상거래 에이전트 설계

백 혜 정[†] · 박 영 택^{††}

요 약

전자상거래 시스템의 보급이 활성화되기 시작하면서 사용자의 구매 행위에 적응형으로 대처하는 지능형 전자상거래 에이전트의 필요성이 증대되고 있다. 이와같은 적응형 전자상거래 에이전트는 사용자의 구매 행위를 모니터링하면서, 각 분야별 고객의 구매 행위를 자동 분류하고, 분류된 각 클러스터로부터 사용자의 취향을 학습하는 하는 기능을 필요로 한다. 이러한 기능을 가지는 적응형 전자상거래 에이전트를 구축하기 위해서 본 논문에서는 다음 3가지 부분에 중점을 두고 시스템을 설계하였다. 첫째, 사용자의 구매 행위를 포괄적으로 모니터링하여 사용자 행위로 추상화하는 모니터 에이전트, 둘째, 고객 구매 행위 데이터로부터 유사한 분야 구매 데이터들로 클러스터하는 개념적 클러스터 에이전트, 셋째, 각 클러스터로부터 사용자 프로파일을 구축하는 사용자 프로파일 에이전트를 중심으로 설계하는 방안을 제안하고 있다. 특히, 본 논문에서는 보다 정확한 고객 구매 행위를 학습하기 위해서 개념적 클러스터링 방식과 귀납적 기계학습 방식을 적용하는 2단계 구조를 제안하고 있다. 이와 같은 구조는 여러 분야의 상품을 구매한 정보로부터 사용자의 다중 취향을 학습할 때 발생하는 문제를 해결함으로써, 사용자 프로파일을 정확하게 구축할 수 있는 장점이 있다. 이러한 정확한 사용자 프로파일을 기반으로 사용자에게 보다 적절한 정보를 제공하는 적응형 전자상거래 시스템을 만들 수 있다.

Design of Adaptive Electronic Commerce Agents Using Machine Learning Techniques

Hey-Jung Baek[†] · Young-Tack Park^{††}

ABSTRACT

As electronic commerce systems have been widely used, the necessity of adaptive e-commerce agent systems has been increased. These kinds of agents can monitor customer's purchasing behaviors, cluster them in similar categories, and induce customer's preference from each category. In order to implement our adaptive e-commerce agent system, we focus on following 3 components - the monitor agent which can monitor customer's browsing/purchasing data and abstract them, the conceptual cluster agent which cluster customer's abstract data, and the customer profile agent which generate profile from cluster. In order to infer more accurate customer's preference, we propose a 2 layered structure consisting of conceptual cluster and inductive profile generator. Many systems have been suffered from errors in deriving user profiles by using a single structure. However, our proposed 2 layered structure enables us to improve the quality of user profile by clustering user purchasing behavior in advance. This approach enables us to build more user adaptive e-commerce system according to user purchasing behavior.

키워드 : 적응형 전자상거래 에이전트(Adaptive Electronic Commerce Agent), 고객 프로파일(User Profile), 개념적 클러스터링(Conceptual Clustering), 귀납적 기계학습(Inductive Machine Learning)

1. 서 론

전자상거래 에이전트는 사용자에게 보다 편리한 기능을 제공하도록 한다. 여러 가지 측면에서 사용자에게 편리한 기능을 제공할 수 있지만, 가장 대표적인 기능은 사용자의 구매 행위에 적응형으로 대처할 수 있는 능력이다. 이와같은 기능은 보다 정확한 사용자 프로파일을 구축하고, 이와

같은 사용자 프로파일을 기반으로 사용자에게 보다 필요한 상품 정보를 제공할 수 있도록 한다. 본 논문에서는 사용자의 구매 행위에 따라서 적절한 상품 정보를 제공할 수 있는 적응형 전자상거래 에이전트를 구축하는 방안을 제안하고 있다.

적응형 에이전트를 구축하기 위해서 본 논문에서는 3가지 부분에 중점을 두고 시스템을 개발하였다. 사용자의 정확한 프로파일을 구축하기 위해서, 사용자의 구매 행위를 모니터링하는 모니터 에이전트, 사용자 구매 행위 데이터로부터 유사한 분야 구매 데이터들로 클러스터하는 개념적 클러스터 에이전트, 각 클러스터로부터 사용자 프로파일을 구축하

※ 이 논문은 1999년도 한국학술진흥재단의 연구비에 의하여 지원 되었음 (KRF 99 041 E00267).

† 준 회 원 : 송실대학교 대학원 컴퓨터학과

†† 정 회 원 : 송실대학교 컴퓨터학과 교수

논문접수 : 2002년 6월 14일, 심사완료 : 2002년 11월 21일

는 사용자 프로파일 에이전트를 구축하기 위한 방안을 제안하고 있다.

이와같은 방식을 취한 이유는 기존의 사용자 프로파일 에이전트가 가지고 있는 문제점을 극복하기 위함이다. 기존의 기계학습 방식을 활용하는 시스템들에서는 입력 데이터의 이질적인 구조로 인해서 구축된 사용자 프로파일이 부정확한 경우가 많이 있다. 예를 들면, 의류를 구매한 데이터와 식료품을 구매한 데이터를 같이 가지고 사용자 프로파일을 구축하면 그 결과는 매우 부정확한 결과를 보이게 된다. 이와같은 문제점을 극복하기 위해서 본 논문에서는 사용자 구매 행위를 보다 정확히 클러스터링하고, 그 클러스터 결과에 귀납적 기계학습을 적용하여 사용자 프로파일을 구축하는 방안을 제안하고 있다.

본 논문에서는 비감독 기계학습과 감독기계학습을 혼합하므로써 사용자가 자신의 관심 분야를 일일이 지정하지 않아도 되는 장점을 가지는 방안을 제안하고 있다. 비감독 기계학습은 기존의 개념적 클러스터링 방식을 개선하는 방식을 활용하고 있다. 기존의 개념적 클러스터링 방식은 하향식(top-down)방식을 취하기 때문에 입력되는 데이터의 순서에 따라서 클러스터에 오류가 발생할 수 있게 된다. 이와같은 오류는 최종 클러스터의 정확도를 떨어뜨리게 된다. 따라서 본 논문에서는 먼저 각 데이터 사이의 유사도를 계산하여 초기 seed를 결정하고 이에 개념적 클러스터를 적용하는 방식을 취하였다. 이와같은 방식은 최종 결과 클러스터들의 오류를 최소화하는 데 기여하고 있다.

본 논문은 2장에 관련 연구를 기술하고 있으며, 3장에서는 적응형 전자상거래 에이전트의 제안된 구조를 설명하고, 4장에는 모니터 에이전트, 5장에는 개념적 클러스터 에이전트, 6장에는 사용자 프로파일 에이전트에 대해서 기술한다. 또한, 7장은 본 논문에서 제안한 적응형 전자상거래 에이전트 구조의 핵심인 개념적 클러스터의 실험 결과와 귀납적 기계학습을 이용한 사용자 프로파일 구축 실험 결과를 설명하도록 한다.

2. 관련 연구

인터넷 기반의 전자상거래 시스템이 급속히 보급되면서 전자상거래 에이전트에 대한 연구가 활발히 진행되고 있다. 전자상거래 에이전트 연구에는 Amazon, BarnesAndNoble, ZD Net, MyLaunch, PersonaLogic과 같은 추천중심의 전자상거래 에이전트와, Kasbah, AuctionBot와 같은 협상중심 전자상거래 에이전트과, Bargain Finder, Jango와 같은 비교 쇼핑 전자상거래 에이전트가 있다[1]. 추천중심의 전자상거래 에이전트는 쇼핑물을 검색하여 사용자에게 그 결과를 추천하는 방식을 취하며, 협상중심 전자상거래 에이전트는 사용자가 제시한 여러 구매 조건을 만족시키기 위해서 구매 에이전트와 판매 에이전트 사이의 협상과정을 거쳐서 추천하는 방식을 취하고 있다[2, 7, 14].

기존의 인터넷 쇼핑물에서는 사용자가 원하는 상품을 검

색하기 위해서 많은 시간이 소요되기 때문에, 이를 개선하기 위해서 BarnesAndNoble, XDNet, MyLaunch, PersonLogic 등에서는 시스템에 지능형을 추구 하고 있다. 특히, 대표적인 추천기반 전자상거래 시스템인 Amazon같은 경우는 사용자가 특정 도서에 관심을 가지고 있을 때, 관심을 가지는 도서를 구입한 다른 사람이 동시에 구입한 도서를 사용자에게 추천하고 있다[8, 9]. 이는 사용자의 현재 관심을 기반으로 하는 단기간의 추천이며, 일반적인 경향 분석에 대한 추천으로, 사용자의 세부적인 관심을 고려할 수 없으며, 모든 사람에게 동일한 추천을 수행할 수 밖에 없다. 그래서 본 연구에서는 이러한 단점을 개선하고자, 사용자의 관심에 대한 지속적인 정확한 학습을 통하여, 개개인에 맞는 서비스를 수행하는 전자상거래 에이전트 시스템을 개발하고자 한다.

또한, 기존의 전자상거래 시스템들은 사용자가 선호하는 상품 군(cluster)에 대한 고려를 정확히 하지 못하는 단점을 가지고 있다. 따라서, 사용자가 브라우징한 여러 상품 군을 입력 데이터로 처리하여 사용자 프로파일을 구축하므로 학습한 사용자의 프로파일에 많은 오류가 존재하게 된다. 본 연구는 이러한 시스템의 문제점을 해결하기 위하여 2단계의 사용자 프로파일을 기반으로 한 전자상거래 에이전트를 제안하였다.

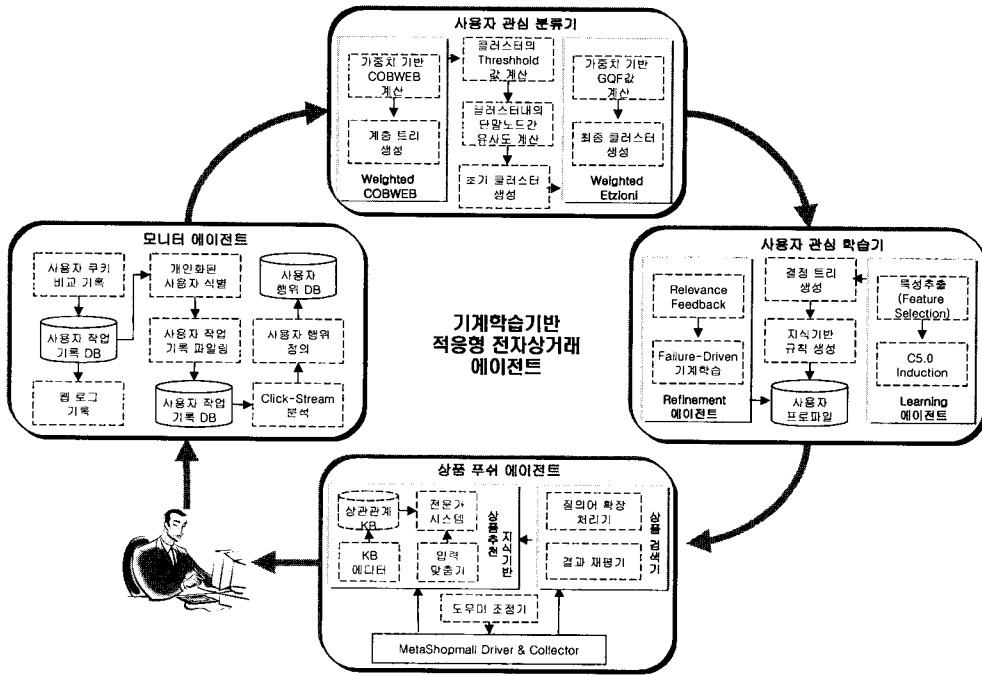
3. 기계학습 기반 전자상거래 에이전트

본 논문에서는 사용자의 행위에 따라서 적응형으로 동작하는 지능형 전자상거래 에이전트 구축을 위한 방안을 제시한다. 지능형 전자상거래 에이전트의 핵심은 전자상거래를 이용하는 사용자의 행위를 정확하게 모니터링하고, 사용자의 의도를 파악하여, 사용자가 관심을 가지는 정보를 추출하여 생성하는 사용자 프로파일의 정확도에 있다. 본 논문에서는 정확한 사용자 프로파일을 학습하고, 이를 기반으로 적응형 서비스를 수행할 수 있도록 다음과 같은 전자상거래 에이전트 구조를 제안한다.

본 논문에서 제안하는 기계학습 기반 적응형 전자상거래 에이전트는 사용자의 행위를 포괄적으로 모니터링하여 사용자 의도를 파악하는 모니터 에이전트, 사용자의 행위에서 관심분야별로 클러스터하는 개념적 클러스터 에이전트, 관심분야를 기반으로 귀납적 기계학습을 통한 사용자 프로파일 생성 에이전트, 지식기반 푸쉬 에이전트로 구성하였다.

첫째, 모니터 에이전트는 전자상거래를 이용하는 사용자의 행위를 포괄적으로 모니터링하여, 사용자의 행위를 추상화한다. 이를 위해, 사용자의 구매 행위를 모니터링하여 각 사용자가 브라우징한 상품, 구매한 상품 등을 로그 파일로 만든다. 이와같은 로그 파일은 각 사용자가 전자상거래 시스템을 브라우징하면서 수행한 행위들을 정확히 기록하는 기능을 수행한다. 이렇게 생성한 로그 파일을 기반으로 사용자의 행위를 추상화하는 기능을 포함한다.

둘째, 사용자 관심분야 클러스터 에이전트는 모니터 에이전트에서 생성한 사용자의 행위로부터 유사한 관심분야별로



클러스터를 만드는 기능을 수행한다. 클러스터 에이전트의 정확성을 극대화하기 위해서 본 논문에서는 하향식(top-down)방식과 상향식(bottom-up)방식을 병행하고 있다. 이와 같은 이유는 기존의 개념적 클러스터 방식은 하향식(top-down)방식을 취하고 있으므로 데이터 입력 순서에 따라서 유사한 상품이 다른 클러스터에 위치하는 경우가 존재하게 된다. 이를 해결하기 위하여, 하향식(top-down)방식과 상향식(bottom-up)방식을 병행하는데, 상향식(bottom-up)방식을 적용하기 이전에 하향식(top-down)방식으로 어느 정도 유사한 상품들은 작은 seed로 만들어 주고 이러한 seed들을 대상으로 상향식(bottom-up)방식을 적용함으로써 유사한 상품은 같은 클러스터에 위치하도록 하는 방안을 제안하고 있다.

셋째, 사용자 프로파일 생성 에이전트는 클러스터 에이전트가 만들어준 각 클러스터에 귀납적 기계학습 방식을 적용하여 사용자의 취향을 학습하는 기능을 수행한다. 본 논문에서는 엔트로피 개념을 활용하는 C5.0 방식을 제안하고 있다. 클러스터 에이전트가 사용자의 관심 상품 분야별로 클러스터를 생성해 주었으므로 각 클러스터에 귀납적 방식

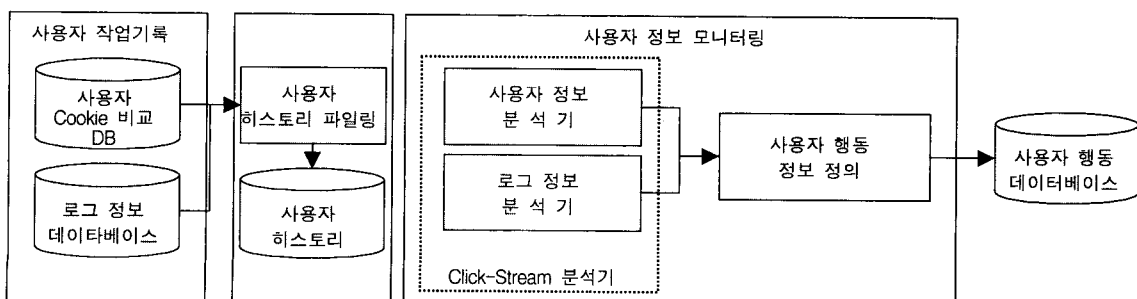
을 적용함으로써 사용자의 프로파일을 보다 정확하게 구축할 수 있게 된다.

마지막으로 상품 정보 푸쉬 에이전트는 학습된 사용자 프로파일을 바탕으로 관련있는 상품을 제안해 주는 기능을 수행한다. 본 논문에서는 정확한 사용자의 사용자 프로파일을 기반으로 유사한 상품이 발생하는 경우 지식기반 방식으로 추천하는 기법을 제안하고 있다.

4. 모니터 에이전트

적응형 전자상거래 시스템을 위해서 가장먼저 수행 되어야 하는 것은 사용자가 취하는 행동들을 포괄적으로 탐지하는 것이다. 그래서 본 논문에서는 모니터 에이전트를 이용하여 사용자의 사용자의 행위를 모니터링하고 각 사용자가 브라우징한 상품, 구매한 상품 등을 로그 파일로 만든후, 분석한다.

본 논문에서는 인터넷을 이용한 사용자에 대한 모든 행위를 탐지하여 저장하는 방법으로 쿠키(Cookie)와 사용자의 웹 로그(Web Log)를 사용하였다. 쿠키(Cookie)는 일반적인



로 웹 페이지를 통해서 사용자 컴퓨터에 매우 작은 텍스트 파일로 저장되는 형태이다. 이렇게 저장된 쿠키를 통해서 사용자 접근했던 웹 페이지에서의 행동에 대한 개인적인 정보들을 저장할 수 있다. 또한 시스템에서는 사용자 쿠키를 가진 사용자가 들어올 경우 사용자 정보 변화에 대한 탐지를 수행하는 중요한 역할을 수행한다. 웹 로그는 사용자가 접속한 웹 서버에서 행위한 모든 작업이 시간별로 저장되어 무슨 작업을 했는지 알 수 있도록 한 것이다. 본 연구에서는 웹 페이지에서 사용자의 행위를 기록하는 방법으로 웹 서버에서 제공하는 로그 파일의 분석 방법과 사용자를 위한 웹 페이지를 통해서 기록하여 분석하는 방법을 병행하여 사용자의 행위 데이터 수집, 기록을 수행한다.

쿠키 및 웹 로그를 통해서 얻어진 정보는 모니터가 쉽게 사용자에게 대한 형식화된 기록 정보를 구축하기 위해서 정보 중에서 필요한 정보와 불필요한 정보를 골라내어 저장한다. 이렇게 정제된 정보를 통해서 정보를 이용하는 모니터 쪽에 불필요한 작업을 제거 할 수 있다. 따라서, 쿠키의 정보와 웹 로그의 정보를 일정 형태로 수정하여 사용자가 행한 기록을 저장한다. 이렇게 모니터된 정보는 사용자별로 프로필을 만들기 위해서 사용자별로 나눈후, 각각의 모니터된 정보를 사용자 행위로 추상화하는 과정을 수행한다. 예를들어, 사용자가 특정 상품 URL에 방문했다는 정보를 특정 상품과, 그 URL에 대한 관심 정보로 추상화 한다. 추천 상품에 대하여 사용자가 일부의 상품에 대한 URL에 대한 방문 기록은 사용자에게 추천 상품에 대한 피드백으로 분석하여, 추천 상품에 대한 사용자의 관심 정도로 정의한다. 이처럼, 본 논문에서 모니터 에이전트는 전자상거래에서 이루어지는 모든 이벤트들을 기록하는 것 뿐 아니라, 사용자 관심여부로 추상화하는 과정을 포함한다.

5. 사용자 관심 분야 클러스터 에이전트

본 논문에서는 보다 정확한 사용자 프로파일을 구축하기 위하여 2단계 구조를 제안한다. 즉, 사용자의 행위 데이터로부터 유사한 관심 분야별로 묶는 클러스터링을 수행한후, 클러스터된 관심 정보를 기반으로 사용자 프로파일을 생성하는 것이다.

사용자 프로파일을 생성하기 위한 1단계인 사용자 관심 분야 클러스터 에이전트는 모니터 에이전트에서 생성한 사용자의 행위로부터 유사한 관심 분야별로 클러스터를 만드는 기능을 수행한다. 본 논문에서는 기존의 클러스터링 방식이 나타내는 입력 의존적인 문제를 효율적으로 완화시키고, 클러스터의 정확성을 극대화하기 위해서 중요 속성의 가중치를 고려한 가중치 기반 통합 방식의 클러스터링을 제안한다. 여기서 통합방식은 하향식(top-down)방식과 상향식(bottom-up)방식을 병행하는 것으로, 하향식(top-down)방식으로 어느 정도 유사한 상품들은 작은 seed로 만들어 주고 이러한 seed들을 대상으로 상향식(bottom-up)방식을 적용함

으로, 유사한 상품은 같은 클러스터에 위치하도록 한다. 본 논문에서는 하향식 방식으로는 COBWEB 클러스터링을 사용하며, 상향식 방식으로는 Etzioni 클러스터링을 이용한다. 본 논문에서는 속성의 가중치를 고려하기 위하여 기존의 COBWEB과 Etzioni의 평가 함수(Evaluation Function)를 수정 제안한다. 가중치 기반의 평가함수를 이용하여 상품정보를 클러스터링하는 방법은 다음과 같다. 첫째, Weighted COBWEB 클러스터링을 적용하여 계층적 구조를 생성하고, 둘째, 초기 클러스터 생성 평가함수를 이용하여 초기 클러스터를 생성하며, 마지막으로 Weighted Etzioni을 이용하여 초기 클러스터 병합통한 최종 클러스터를 생성한다.

다음은 본 논문에서 제안하고 있는 상품의 속성의 가중치를 이용한 Weighted Cobweb 클러스터링과 Weighted Etzioni 클러스터링을 위한 평가 함수에 대해 설명하도록 한다.

5.1 Weighted COBWEB 클러스터링

COBWEB은 원래 인간의 점진적인 개념학습을 모델링 하기 위해 개발된것으로서, 데이터들의 상호 연관성을 이용하여 개념을 형성해 가고 형성된 개념을 이용하여 새로운 예제를 분류한다. COBWEB은 상호 연관성의 측정 값으로 Category Utility[6]를 기반으로 하는 평가 함수를 사용한다. 이때, 평가 함수는 한 데이터의 속성의 값이 주어졌을때, 특정 클러스터에 속할 확률을 의미하는 Predictiveness과, 한 데이터가 특정 클러스터일때, 속성이 특정값을 가질 확률을 의미하는 Predictability의 곱으로 계산한다[4, 5]. 이를 식으로 나타내면 식 (1)과 같다.

$$\sum_k \sum_i \sum_j P(C_k | A_i = V_{ij}) P(A_i = V_{ij} | C_k) \tag{1}$$

위에서 $P(C_k)$ 란 클러스터 k 의 전체에 대한 비율이고, $P(A_i = V_{ij} | C_k)$ 는 주어진 클러스터에 대하여 상품 각각의 속성 (A_i)이 특정 값 (V_{ij})을 가지는 확률을 나타내며, i 는 속성의 개수, j 는 학습개체의 개수를 나타낸다. 이때, 각 속성의 값을 $P(A_i = V_{ij})$ 와 각 속성의 가중치 ($W(A_i)$)라 했을 때, 이를 식 (1)에 적용하면 다음과 같은 식을 얻을 수 있다[4, 11].

$$\sum_k \sum_i W(A_i) \sum_j P(A_i = V_{ij}) P(C_k | A_i = V_{ij}) P(A_i = V_{ij} | C_k) \tag{2}$$

식 (2)를 베이저언 정리($P(A_i = V_{ij})P(C_k | A_i = V_{ij}) = P(C_k)P(A_i = V_{ij} | C_k)$)를 이용하면 가중치를 고려한 값은 다음과 같이 정리할 수 있다.

$$\sum_k P(C_k) \sum_i W_i \sum_j P(A_i = V_{ij} | C_k)^2 \tag{3}$$

Gluck and Corter에 의한 Category Utility는 식 (1)에서 계산한 어떤 개체의 속성이 기대되는 기대치의 합(식 (3))에서 분류를 고려하지 않은 기대치값(식 (4))를 뺀값을 부

류의 개수(K)로 나눈 값으로 정의했다[4]. 이때, 속성이 가중치를 가진다고 하였을 때, 분류를 고려하지 않은 기대치 값은 다음과 같다.

$$\sum_{i=1}^I W(A_i) \sum_{j=1}^I P(A_i = V_{ij})^2 \quad (4)$$

결과적으로, 본 논문에서 제안하는 수정된 Category Utility는 다음과 같다.

$$\frac{\sum_{k=1}^K P(C_k) \sum_{i=1}^I W(A_i) \sum_{j=1}^I [P(A_i = V_{ij} | C_k)]^2}{\frac{\sum_{i=1}^I W(A_i) \sum_{j=1}^I P(A_i = V_{ij})^2}{K}}$$

5.2 Weighted Etzioni 클러스터링

Etzioni의 클러스터링 방식은 클러스터의 응집도를 측정하는 GQF(Global Quality Function)이라는 평가함수를 제안하고 있다[3]. Etzioni의 클러스터링 방식은 클러스터의 질을 정량화하기 위해서 GQF(Global Quality Function)이라는 평가함수를 제안하고 있다[3]. Etzioni의 클러스터링 방식은 매번 반복할때 마다 GQF 함수의 값을 최대화 하는 두 개의 클러스터를 병합한다. 이런 과정은 GQF 함수의 값이 더 이상 증가하지 않을 때까지 반복한다. 이 방식에서는 클러스터 c의 응집도를 클러스터에 속하는 모든 것들에 공통적으로 나타나는 속성의 수로 정의하고 h(c)로 표시한다. 그리고 단일 클러스터의 스코어는 클러스터를 구성하는 상품의 수와 정규화된 클러스터 응집도의 곱으로 정의하고 s(c)로 표시한다. 아래의 수식은 클러스터 스코어를 계산하는 수식이다.

$$s(c) = |c| \times \frac{1 - e^{-\beta h(c)}}{1 + e^{-\beta h(c)}}$$

Weighted Feature의 중요도를 고려하기 위해서 본논문은 h(c)를 다음과 같이 정의한다.

$$h(c) = \sum_i W_i \times T_i$$

$T_i = 1$ (If 클래스내에 attribute i의 Value가 같다)

$T_i = 0$ (If 클래스내에 attribute i의 Value가 같지 않다)

즉, h(c)는 클러스터 c의 응집도를 클러스터에 속하는 모든 데이터들에 공통적으로 나타나는 각 속성(feature)의 중요도 합으로 계산한다. 이때, β 는 클러스터의 크기와 클러스터의 응집도 사이의 Trade-off를 정하기 위한 변수이다. β 값은 경험적으로 0에서 1사이의 값을 가지는 것이 바람직하다. 이때, 전체 클러스터 C에 대해서 GQF(C)를 구하는 수식은 다음과 같다.

$$GQF(C) = \frac{f(C)}{g(|C|)} \sum_{c \in C} s(c)$$

GQF(C) 함수는 크게 세부분으로 이루어진다. 먼저 f(C)는 전체 집합중에서 개수가 두 개 이상인 클러스터에 포함되는 비율이다. 다음으로 g(|C|)는 클러스터의 개수를 표시한다. 그리고 s(c)의 합은 모든 클러스터 스코어의 합을 나타낸다. GQF 평가함수의 1/g(|C|) 항과 $\sum_{c \in C} s(c)$ 항은 낮은 응집도의 소수의 클러스터를 생성하려는 경향과 높은 응집도의 다수의 클러스터를 생성하려는 경향을 각각 생성하고 GQF 함수는 적절한 Trade-off 지점을 찾아준다.

기존의 클러스터링 수행을 정지하는 임의의 임계값을 가졌던 반면에 GQF 함수는 자체 수식 내에서 클러스터 생성을 정지할 시점을 찾아준다. 또한 클러스터의 응집도를 클러스터 내의 상품의 속성으로 정의함으로써 클러스터링 결과가 관련이 적은 상품들이 포함되는 것을 방지할 수 있다.

6. 사용자 프로파일 생성 에이전트

사용자 프로파일 생성 에이전트는 사용자 프로파일 구축을 위한 2단계로서, 1단계의 사용자 관심 클러스터를 기반으로 사용자의 프로파일을 구축하고, 사용자의 관심을 학습을 한다. 본 논문에서는 사용자의 관심을 학습하기 위해서 엔트로피 개념을 활용하는 C5.0이라는 귀납적 기계학습 알고리즘을 이용한다. 본 논문에서 사용하는 C5.0은 Ross Quinlan이 분류모델(Classification Model)로서, 클러스터를 대상으로 각 클러스터를 대표하는 특성(feature)를 발견하고 분석한다[12, 13]. 분석된 정보는 영역을 분류하는 모델을 구성하고 이 모델은 각각의 속성값에 따라 결정 트리를 생성한다. 이때 C5.0은 결정 트리를 만들기 위해서 정보 이론(Information Theory)에 근거하는 gain값을 사용하는데 gain값을 구하는 식은 다음과 같다[16, 17]. 여기서 S는 전체 집합이며, A는 속성을 나타내고, S_v는 속성의 값을 나타낸다.

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

C5.0을 이용하여 생성되는 결정 트리는 결정 변수를 나타내는 내부 노드와 분류결과를 나타내는 단말노드 그리고 이들을 서로 연결하는 선(edge)으로 이루어져 있다. 이러한 트리의 생성은 사용자 관심 정보를 파악하는데 중요한 속성순으로 생성되며, 지식의 형태로 인식되어 지식 베이스에 저장한다. 즉, C5.0은 결정 트리를 이용하여 중요한 속성에 대한 일정한 규칙을 생성하고, 이 규칙들은 키워드에 따른 카테고리의 분류를 가능하게 할 뿐만 아니라 카테고리에 따른 속성의 추출도 가능하다.

본 논문에서는 사용자 관심 학습전에 사용자의 관심 상품 분야별로 클러스터를 생성을 수행함으로, 귀납적 학습의 입력을 균일(homogeneous)하게 유지하여, 사용자의 관심 프

로파일을 보다 정확하게 구축할 수 있다. 그리고, 본 논문에서 많은 학습 중에서 귀납적 기계학습을 이용하는 이유는 사용자와의 상호작용을 하면 할수록 이전보다 더욱 효과적인 학습을 처리하는 적응형 시스템이 가능하기 때문이다.

일반적으로 귀납적 기계학습을 이용하여 생성된 내용은 일반적으로 귀납적 기계학습(inductive learning)에 의해서 학습된 내용은 불완전(incomplete), 부정확(incorrect)한 특성을 가지게 된다[13]. 따라서, 이러한 문제를 보완하여 보다 완벽하고 정확한 사용자 프로파일을 구축하기 위해서 본 논문에서는 failure-driven 기계 학습 방식을 이용한다[10, 15]. 이는 사용자 프로파일을 이용하여 추천한 상품에 대한 사용자의 피드백을 분석하여 실패(failure)로 간주되는 경우를 이용하여 사용자 프로파일에 반영하는 것이다. 에이전트가 제시한 상품 정보를 처리하는 사용자의 행위에 따라서 다음과 같은 종류를 실패(failure)로 정의한다. 첫째, 에이전트가 높은 확신(confidence)도를 기반으로 제시한 상품을 사용자가 관심 없어 하는 경우와 둘째, 에이전트가 낮은 확신(confidence)도를 기반으로 제시한 상품을 사용자가 관심을 가지고 구매하는 경우이다. 본 논문에서는 사용자의 관심도를 보다 정확하게 학습하기 위하여, 실패(failure)가 발생한 경우 사용자 프로파일의 불완전하고, 부정확한 부분을 알아내어 정제(refinement)한다. 이를 위해서, 사용자에게 정보를 제공하게 된 추론 과정을 의존 구조(dependency structure)로 구축하고 이를 근거로 관련이 있는 사용자 프로파일을 알아내고 이를 사용자에게 수정할 수 있도록 한다.

이처럼 학습된 사용자 프로파일은 지식기반 상품 푸쉬 에이전트를 통하여 사용자에게 필요한 상품을 제안하는 기능을 수행한다. 본 논문에서는 정확한 사용자의 사용자 프로파일을 기반으로 유사한 상품이 발생하는 경우 지식기반 방식으로 추천하는 기법을 제안하고 있다. 상품 추천 에이전트는 일반적인 상품 연관 규칙과 사용자 프로파일을 기반으로 하는 사용자마다의 독특한 연관 규칙을 지식베이스로 구축한다. 이렇게 구축된 지식 베이스는 전문가 시스템 기법을 활용하여 사용자에게 관심 정보를 추천한다. 사용자의 모니터 정보를 대상으로 학습된 사용자 프로파일을 구성하는 중요 속성들에 대한 전자상거래내의 전문가들의 지식을 입력하여 이에 대한 정보를 근거로 하여 사용자에게 자신의 관심 정보에 의한 추천뿐만 아니라 전문가의 지식을 추가한 정보를 추천하게 된다. 다음은 본 논문에서 활용한 지식 베이스내의 규칙의 예이다.

```
rule1000 says if customer(X) and buy(X, Y, Price, T) and
  stimated_next_buy(X, Y, T1) and today(T2) and
  less_than(T2, T1) and sale(Y, T2, Z)
  then advise(buy(X, Y, at(Z))).
```

본 논문에서는 위와 같은 지식 베이스내의 규칙을 이용하여 추천할 상품을 추론 하는데 전문가 시스템을 활용하였다.

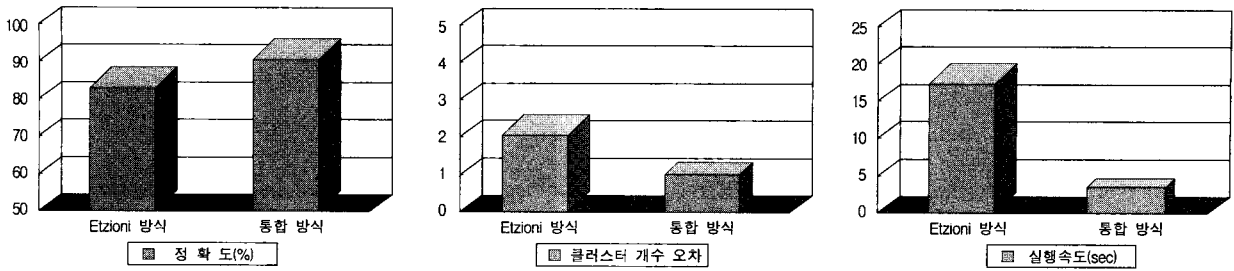
7. 실험

본 논문에서 제안하는 적응형 전자상거래 에이전트는 사용자의 구매 행위에 적응형으로 대처함으로 사용자에게 편안한 기능을 제공한다. 이와같은 기능은 수시로 변하는 사용자의 구매 성향을 파악하고, 이를 기반으로 사용자에게 필요한 상품정보를 제공함으로 이루어진다. 이러한 사용자의 구매 성향의 파악과 필요한 상품 제공은 사용자 프로파일이 정확하게 구축되어야 가능하다. 그래서, 본 논문에서 제안하는 적응형 전자상거래 에이전트의 성능을 평가하기 위하여 사용자 프로파일을 구축의 정확성 평가와 구축된 사용자 프로파일에 대한 추천 정보에 대한 정확도 평가를 수행하도록 한다.

첫째, 사용자 프로파일 구축의 정확성 평가이다. 본 논문에서는 사용자 프로파일을 구축하기 위하여 통합 클러스터링 방식을 수행한 후 이를 기반으로 사용자 프로파일을 구축하였다. 구축된 사용자 프로파일의 정확성을 평가하기 위해서는 먼저, 사용자 프로파일 구축전에 클러스터링을 수행하는 것이 성능 향상에 도움이 되었는지, 그리고 수행된 클러스터링이 정확한지에 대한 실험이 필요하다. 이때, 사용자 프로파일을 수행하기 전에 클러스터링을 수행이 성능 향상에 도움이 된다는 것은 기존 연구에 의해 증명되었다[18]. 그래서, 본 사용자 프로파일 정확성 평가에서는 사용자 프로파일 구축전에 수행되는 클러스터링에 대한 정확도에 대한 실험을 수행한다. 본 논문이 제안하는 클러스터링 방식의 우수성을 평가하기 위하여 기존의 클러스터링 방식과 비교 실험을 수행하였다. 본 실험에 사용하는 데이터는 쇼핑몰에 로그인을 하여 자신이 관심 있어 하는 상품에 대하여 브라우징이나 구매 같은 행위에 대한 저장 내용이다. 사용된 실험 데이터는 총 24개 카테고리에 400여개의 제품 정보를 담고 있다. 각기 다른 사용자 6명에 대한 저장 데이터를 기반으로 클러스터링 실험을 실시하였다.

본 논문에서는 실험의 평가 측도로 정확도의 평균, 평균 클러스터 개수 오차, 평균 실행 속도를 이용한다. 각 실험의 정확도 계산은 전체중에서 최종 클러스터에 맞게 분류된 것의 비율로서 계산하는 평균정확도를 구하고, 평균 클러스터 개수 오차는 실험의 입력 카테고리 개수와 출력 카테고리 개수의 차를 의미하며, 평균 실행 속도는 각 클러스터링이 시작된 시각과 끝난 시각의 차를 계산한다[18]. 다음의 그림은 본 논문에서 제안하는 방식과 기존 방식들에 대한 실험 결과를 평균 정확도 측면, 평균 클러스터 개수 오차 측면, 평균 실행 속도 측면에서 그래프로 표현한 것이다. 표에서 통합 방식은 Weighted Cobweb + Weighted Etzioni 방식을 이용한 클러스터 방식이다.

첫 번째 결과는 클러스터링 방식의 실험 결과를 평균 정확도 측면에 대한 것이다. Etzioni 클러스터링 방식의 평균 정확도는 84%, 통합 클러스터링 방식은 96%의 정확도를 보였다. 정확도 측면에서는 제안하는 방식이 약간 우수한 성



능을 나타나고 있다.

두 번째 결과는 각 클러스터링 방식의 실험 결과를 평균 클러스터 개수 오차 측면에 대한 실험이다. Etzioni 클러스터링 방식의 평균 클러스터 오차는 2.14, 통합 클러스터링 방식은 0.9의 클러스터 개수 오차를 보였다. 평균 클러스터 개수 오차 측면에서는 통합 방식이 다른 방식 보다 더 정확한 클러스터링 결과를 생성하였다. 세 번째는 평균 실행 속도 측면에서 실험한 것이다. Etzioni 클러스터링 방식의 평균 실행 속도는 18.4초, 통합 클러스터링 방식은 2.08초의 실행 속도를 보였다. 평균 실행 속도 측면에서는 통합방식이 Etzioni 방식과 비교해서 약 8배 빠른 클러스터링을 수행하였다.

실험 결과에 따르면, 제안하는 클러스터링 방식은 기존의 클러스터링 방식과 다소 좋은 정확도를 유지하면서 Etzioni의 GQF 함수를 이용하는 클러스터링 방식보다 약 8배 빠른 실행 속도를 보였다. 이때, 클러스터링의 정확도가 다소 높아진 이유는 속성의 가중치를 이용함으로 입력 순서에 따른 의존도를 다소 완화시켜 응집도가 높은 클러스터를 초기 클러스터로 분류해냈으며, 가중치에 근거한 클러스터를 수행함으로써 최종 클러스터의 정확성을 높였기 때문이라고 분석된다. 그리고 실행 속도 측면에서 기존의 방식보다 빠른 이유는 본 논문에서 제안하는 통합방식 클러스터링 방식은 Top-down 클러스터링 방식인 COBWEB의 장점과 Bottom-up 클러스터링 방식인 Etzioni 클러스터링의 장점을 효과적으로 통합하였기 때문이다. COBWEB은 분류 트리를 생성하는 속도가 $O(n \log n)$ 으로 선형적인 알고리즘인 반면 GQF 함수를 이용하는 Etzioni의 방식은 계산 복잡도가 $O(n^2)$ 이다. 그러므로 COBWEB을 이용하여 충분한 응집도와 크기를 가지는 초기 클러스터를 생성한 후에 Etzioni의 방식을 적용하면 각 데이터에서부터 초기 클러스터를 생성할 때와 비교해서 클러스터간 병합 연산을 약 1/8로 줄일 수 있다.

둘째, 구축된 사용자 프로파일에 기반으로 추천 정보에 대한 정확도 향상에 대한 실험을 수행한다. 본 논문에서는 귀납적 기계학습 방식을 적용하여 사용자 프로파일을 생성하게 되며, 이 사용자 프로파일은 사용자 관심 정보 규칙을 가지고 있다. 이 정보는 사용자가 관심을 가지는 상품을 추천하는 지식으로 활용된다. 본 논문에서는 사용자 프로파일을 기반으로 추천한 정보에 대한 성능 평가는 기존의 진단 데이터를 이용하였다. 이는 현재 운용되고 있는 전자 쇼핑

몰에 직접적으로 적용하여 정확성을 판단 할수 없으므로, 간접적으로 진단 데이터에 대한 실험을 통하여 프로파일 기반의 추천 정보의 우수성을 판단하고자 함이다. 본 실험은 237개의 속성과 속성값을 가진 11500 예제에서, 10개의 클래스에서 수행하였다. 수행한 결과, 전문가의 지식을 이용하여 만든 규칙을 이용하여 각 사용자의 정보를 추론한 결과 50%정도의 정확성을 가지고 있었다. 하지만 기계학습을 통하여 구축된 사용자 프로파일 정보를 기반으로 추론하였을 때 80%까지 정확도가 향상되었다. 이러한 실험 결과는 간접적으로 귀납적 기계학습을 통하여 구축한 사용자 프로파일 기반의 추천 정보가 정확하다는 것을 보여주고 있다.

위 두 가지 실험 결과를 통하여, 본 논문에서 제안한 방식이 사용자의 프로파일을 정확히 학습하며 이를 기반으로 하는 추천 시스템이 정확하다는 것을 확인 할 수 있었다. 이러한 실험 결과를 통하여 적응형 전자상거래 에이전트 시스템의 성능 향상을 알 수 있다.

8. 결 론

적응형 전자상거래 에이전트는 사용자의 구매 행위를 모니터링하면서, 각 분야별 사용자의 구매 행위를 자동 분류하고, 분류된 각 클러스터로부터 사용자의 취향을 학습하는 하는 기능을 필요로 한다. 이러한 기능을 가지는 적응형 전자상거래 에이전트를 구축하기 위해서 본 논문에서는 다음 3가지 부분에 중점을 두고 시스템을 설계하였다. 첫째, 사용자의 구매 행위를 포괄적으로 모니터링하여 사용자 행위로 추상화하는 모니터 에이전트, 둘째, 사용자 구매 행위 데이터로부터 유사한 분야 구매 데이터들로 클러스터하는 개념적 클러스터 에이전트, 셋째, 각 클러스터로부터 사용자 프로파일을 구축하는 사용자 프로파일 에이전트를 중심으로 설계하는 방안을 제안하였다. 특히, 본 논문에서는 사용자의 프로파일을 정확히 학습하기 위하여 2단계 구조를 제안하였다. 이러한 2단계 구조는 사용자의 구매 행위를 카테고리 별로 정확히 클러스터링하고, 그 클러스터에 귀납적 기계학습 방식을 적용하는 구조를 제안함으로써 보다 정확한 사용자 프로파일을 구축할 수 있었다. 또한, 사용자의 구매행위를 클러스터링하기 위해서 기존의 하향식(top-down)방식을 보완하여 하향식(top-down)과 상향식(bottom-up)방식을 혼합하는 방안을 제안하였다. 이 방식은 기존의 개념적 클러스터링 방식이 입력 데이터의 순서에 따라서 작은 클러스터가 여

러게 생기고 클러스터의 정확도가 떨어지는 문제점을 극복할 수 있었다. 즉, 입력된 자료를 하향식(top-down)을 적용하여 일정한 크기의 초기 seed를 구축한 후에 그 결과 seeds들에 대해서 개념적 클러스터링 방식을 적용하는 방안을 제안하고 있다. 귀납적 기계학습 방식은 기존의 엔트로피 개념을 이용한 결정트리를 구축할 수 있는 C5.0 방식을 활용함으로써 각 분야별 사용자 프로파일을 정확히 구할 수 있었다. 이와 같이 구축된 각 분야별 사용자 프로파일은 전자상거래 시스템에서 활용되어 사용자에게 유용한 정보를 적시에 제공할 수 있는 적응형 전자상거래 에이전트로서 활용이 가능하게 되었다.

본 논문에서 제안하고 있는 적응형 전자상거래 에이전트는 사용자 개인의 프로파일을 기반으로 시스템을 구성하였으므로, 본 논문에서 제안한 적응형 전자상거래 에이전트 시스템의 우수성 평가는 현재 사용자 프로파일에 대한 정확도를 기반으로 실시하였다. 하지만 보다 정확한 평가를 위해서는 사용자 프로파일을 기반으로 추천시 사용자의 만족도 평가가 병행되어야 한다. 추후 진행되는 연구에서는 사용자 만족도를 기반으로 하는 사용자 프로파일에 대한 정확도를 평가하고자 한다. 이를 위해서 사용자 추천 정보에 정확도와 연관도를 사용자의 추천 정보에 대한 상호 작용을 통하여 정확하게 평가 할 수 방법도 연구 하고자 한다.

참 고 문 헌

[1] BargainFinder. <http://bf.cstar.ac.com/bf/>.
 [2] A. Chavez and P. Maes, Kasbah : An Agent Marketplace for Buying and Selling Goods. Proceedings of PAAM'96, London, UK, pp.75-90, April, 1996.
 [3] Oren Zamir, Oren Etzioni, Omid Madani and Richard M. Karp, "Fast and Intuitive Clustering of Web Documents," Kdd'97.
 [4] Fisher, D. H. and Langley, P., "Methods of conceptual clustering and their relation to numerical taxonomy," In W. Gale(Ed.), Artificial Intelligence and Statistics, Addison Wesley, 1986.
 [5] Gennari, J. H. , Langley, P. and Fisher, D. H., "Models of incremental concept formation," Artificial Intelligence, pp. 11-61, 1989.
 [6] Gluck, M. and Corter, J., "Information, uncertainty and the utility of categories," Proceedings of the Seventh Annual Conference of the Cognitive Science Society, Irvine, CA : Lawrence Erlbaum, pp.283-287, 1985.
 [7] Matthias Klusch, Dynamic Supply Chain Structuring for Electronic Commerce Among Agents, In "Intelligent Information Agents," Chapter 10, Springer, 1999.
 [8] LawrenceHolder, Meral Ozsoyoglu, "Data mining, Data Warehousing & OLAP and ECommerce Breakout Group Report," NSF IDM Workshop, 2001.
 [9] Yazdi Lashkari, Max Metral and Pattie Maes, "Collabora-

tive Interface Agents," Conference of the American Association for Artificial Intelligence, Seattle, August, 1994.
 [10] P.Maes, Intelligent Software. Scientific American, Scientific American, Inc., Vol.273, No.3, pp.84-86, September, 1995.
 [11] Kathleen Mckusick, Kevin Thompson, "COBWEB/3 : A Portable Implementation," NASA Ames Research Center, Technical Report FIA-90-6-18-2, 1990.
 [12] Tom Mitchell, Robert Armstrong, Dayne Freitag and Thorsten Joachims, "WebWatcher : A Learning Apprentice for the World Wide Web," 1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments, Stanford, March, 1995.
 [13] T. M. Mitchell, "Machine Learning," McGraw Hill, 1997.
 [14] Giorgos Zacharia, Alexandros Moukas, Robert Guttman and Pattie Maes, "An agent system for comparative shopping at the point of sale," Proceedings of PAAM'98.
 [15] Ram, A. and Cox, M. T., Failure-driven learning as input bias. In Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society, Hillsdale, NJ : Lawrence Erlbaum Associates, pp.231-236, 1994a.
 [16] R. Quinlan, "Induction of Decision Tree," Machine Learning, pp.81-106, 1986.
 [17] J.R. Quinlan, "C4.5 Programs for Machine Learning," San Mateo, CA : Morgan, Kaufaman, 1993.
 [18] 양찬범, "웹 에이전트를 위한 통합방식 문서 클러스터링", 석사학위논문, 1999.
 [19] Wetschereck, D., Aha, D. W. and Mohri, T. "A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms," Artificial Intelligence Review, pp.273-314, November, 1997.



백 혜 정

e-mail : hjbaek@it.soongsil.ac.kr

1995년 숭실대학교 컴퓨터학과 졸업
(학사)

1998년 숭실대학교 대학원 컴퓨터학과
졸업(공학석사)

1999년~현재 숭실대학교 대학원 컴퓨터
학과 박사과정

관심분야 : 인공지능, 에이전트, 전문가 시스템



박 영 택

e-mail : park@computing.soongsil.ac.kr

1978년 서울대학교 전자공학과 졸업(학사)
1980년 KAIST 전산학 졸업(석사)

1992년 Univ. of Illinois at Urbana-Champaign 졸업(박사)

1981년~현재 숭실대학교 컴퓨터학과 교수

관심분야 : 인공지능, 에이전트, 전문가 시스템