

사용자의 결합된 행동을 이용한 적응형 정보여과 에이전트

송용수⁰ 홍언주 오경환
한국관광대학 디지털콘텐츠과, 서강대학교 컴퓨터학과
digital7@sogang.ac.kr, hej@kctc.ac.kr, kwoh@ccs.sogang.ac.kr

An Adaptive Information Filtering Agent based on User' s Combined Behaviors

Yong-Soo Song⁰ Eon-Joo Hong Kyung-Whan Oh
Dept.of Digital Contents Korea Tourism College
Dept. of Computer Science Sogang University,

요 약

본 논문에서는 온라인 뉴스 기사를 여과하여 사용자에게 관련있는 뉴스기사만을 선별적으로 여과하여 보여주는 정보여과 에이전트를 설계 및 구현하였다. 정보여과의 핵심이라고 할 수 있는 정확한 사용자 프로파일 구축과 정보에 대한 사용자의 적합성 반응인 명시적 피드백과 암시적 피드백을 모두 결합한 피드백을 사용하여 사용자 프로파일을 좀 더 정교하게 구축하는 방법을 기술하였다. 실험을 통하여 사용자의 결합된 적합성 피드백 행동에 기반한 정보여과 에이전트의 성능이 단일의 피드백만을 사용했을 때보다 더 좋은 정확성과 적응성을 지니고 있음을 보여 주었다.

1. 서 론

컴퓨터기술의 급속한 발달과 인터넷의 빠른 보급과 확산으로 인하여 정보 근원의 매체는 더 추가되었고, 비용이 매우 저렴한 대용량의 저장 장치 덕분에, 오늘날 현대인에게 쏟아져 나오는 지식 정보의 양은 그 만큼 비례하여 증가하게 되었다. 이런 정보의 과부하 속에서 자신에게 관련 있는 정보와 없는 정보를 분류, 구별하느라 많은 시간과 노력을 기울이고 있는 실정이다. 유용한 정보찾기에 쏟아 부는 여력이 크기 때문에 상대적으로 적은 시간과 노력만으로도 최적의 정보를 찾을 수 있는 방법과 대안을 찾으려 하였고 이와 같은 현실적 문제를 해결하기 위하여 등장한 컴퓨터 기반 응용 기술이 바로 정보여과기법을 지닌 지능형 에이전트이다. 이는 사용자에게 과도하게 전달되어지는 많은 정보 속에서 관련없는 정보는 걸러내고, 사용자에게 관련있는 유용한 정보만을 선별하여 보여주는 시스템인데, 기존 사용자가 직접적으로 해왔던 업무를 대신 해줌으로써 많은 시간과 노력을 절감시켜 줄 수 있는 편리함과 효율성의 기능을 지니고 있다. 본 논문에서 개인화된 정보여과 뉴스 에이전트를 구축함과 동시에 보다 효율적으로 사용자 관심분야에 관련된 문서들을 여과하기 위하여 사용자의 적합성 반응인 결합된 피드백을 이용하는 방법을 기술하고자 한다. 뉴스 에이전트가 효율성 높은 정보여과를 이루기 위한 핵심인 사용자 프로파일 구축에 있어, 필수 요건인 사용자의 적합성 피드백으로 명시적 피드백을 주 적합성 피드백으로 삼고 동시에 암시적 피드백을 보조적인 적합성 피드백으로 활용하려고 하였다. 이 두 가지 피드백 행위들을 적절히 결합하여 사용자의 문서에 대한 좀 더 적절한 적합성 평가와 반응을 얻으려 하였다.

2. 연구 배경

2.1 정보 여과

정보여과는 최근 거대한 정보의 흐름 속에서 많은 정보를 다루기 위해 빠르게 발전하고 있는 분야중의 하나이다. 정보여과의 목적은 끊임없이 들어오고 있는 데이터의 흐름으로부터 관련없는 데이터는 제거하고 사용자에게 관련있는 유용한 정보만을 보여줌으로써 시간과 노력을 감소시키려는 데 목적이 있다. 오늘날 실세계 많은 정보여과 시스템들이 다양한 영역 속에서 개발되어져 왔고, 정보여과 시스템은 구조화되지 않은, 혹은 반구조화된 많은 양의 데이터를 다루고 있으며, 사용자 프로파일 에 기반하여 여과를 실시하고 있는 것이다. 정보여과의 여러 응용분야에서 공통적인 중요한 핵심은 바로 정보여과를 원하는 사용자 프로파일을 정확하게 구축하고, 사용자에게 펼쳐진 정보에 대한 적절한 행동과 반응을 인식하는 것이다.

2.2 사용자 모델링

사용자 모델링이란 어떤 시스템이 자신의 목적을 보다 효과적으로 수행할 목적으로 사용자의 독특한 행동들을 특화하면서 파악해 나가는 과정을 말한다. 사용자 모델을 구축하는 형태는 모델 구축시 사용자의 의식적인 개입 여부에 따라 암시적 사용자 피드백에 바탕을 둔 사용자 모델링과 명시적 사용자 피드백에 바탕을 둔 사용자 모델링으로 구분할 수 있다. 암시적 사용자 모델링이란 사용자 모델의 구축과정에서 사용자의 의식적인 개입이 이루어지지 않는 것을 말한다. 명시적 사용자 모델링에 비교하여 Accuracy는 떨어지나 Coverage는 우수하다. 명시적 사용자 모델에서는 사용자 모델의 구축과정에서 필수적으로 사용자의 의식적인 개입이 이루어진다. 시스템이 사용자에게 특정 문서에 대하여 구체적이고 정확한 관심도 표현을 요구하는 방법이므로, 이를 통해 시스템은 보다 정확한 프로파일의 학습과 효과적인 관심 분야의 문서를 보여줄 수 있는 장점이 있지만

사용자에게 문서에 대한 인지심리적인 측면에서 부담을 준다는 단점이 있다. 본 논문에서는 각 모델링의 단점을 줄이고 장점작성자: 김성관을 취합하여 더욱 정교하고 구체적인 사용자 결합된 피드백에 바탕을 둔 사용자 프로파일을 구축하려는데 그 목적이 있다. 명시적 모델링의 장점이라고 할 수 있는 Accuracy와 암시적 모델링의 장점이라고 할 수 있는 Coverage 모두를 얻어 전체적인 정보여과의 Precision을 얻으려는 시도이다.

제 3장. 설계 및 구현.

본 논문에서 구현한 정보여과 뉴스 에이전트는 영문 텍스트 문서(유즈넷 뉴스 기사)를 대상으로 하였으며, 2단계로 이루어 진다. 첫번째 단계는 문서와 사용자 프로파일을 벡터공간으로 표현하는 과정과 프로파일과 문서들의 유사성을 비교하여 일정 기준에 의해 사용자에게 제공될 문서를 선택하는 과정이고, 두번째 단계는 사용자의 명시적 피드백과 암시적 피드백을 결합한 적합성 피드백 행위들을 기반으로 프로파일을 학습시키는 과정이다. 사용자 프로파일 학습과정은 프로파일 내의 가중치를 변화시키는 것으로 생각할 수 있다. 특정 분야에 대한 사용자의 관심의 증가는 해당 키워드의 추가 또는 이미 존재하는 키워드의 가중치 증가를, 사용자의 관심의 감소는 해당 키워드의 가중치 감소 또는 해당 키워드의 삭제로 나타난다. 이와 같은 키워드의 추가, 삭제는 사용자 특성의 변화에 적용할 수 있도록 한다. 본 논문에서 제안하는 사용자 프로파일 학습은 사용자의 결합된 피드백 행위인 암시적 피드백 행위와 명시적 피드백 행위에 기반하여 이루어진다. 다음은 각각의 피드백 행위에 대한 가중치 부여를 나타낸 표이다.

<표 1> 사용자의 암시적 피드백 행위

1	Print	1
2	Save	1
3	Read	0.5
4	Delete	-1

<표 2> 사용자의 명시적 피드백 행위

1	High Relevant	1
2	Relevant	0.5
3	Normal	0
4	Irrelevant	-0.5
5	High Irrelevant	-1

본 논문에서 제안하는 사용자 프로파일 학습은 사용자의 결합된 피드백에 기반하여 이루어진다. 이와 같은 학습과정은 다음의 식 같이 표현될 수 있다.

$$P_{new} = P_{old} + \alpha \times f_{Combined} \times D$$

P,D는 각각 사용자 프로파일과 문서를 나타내는 벡터 표현이며, α 는 학습률을 나타내며, $f_{Combined}$ 는 사용자의 결합된 피드백을 나타낸다. 위에 각각 설정한 암시적, 명시적 피드백의 가중치를 이용하여 다음 식에 나타내는 방법으로 결합된 피드백을 설정하도록 한다.

$$f_{Combined} = \sum_{j \in F_j} f_{ij} + \sum_{k \in F_k} f_{ek}$$

위 식을 적용하는 과정에서 추가적인 규칙을 따르도록 한다. 암시적 피드백 행위에서 인쇄, 저장 행위가 동시에 각각 2회이상 발생할 경우 2이상의 가중치를 초과할 수 없도록 하며, 사용자가 선택하지 않아 문서를 읽지 않았을 경우에는 -1의 값으로 설정하여 프로파일 내의 가중치가 하향 조정 되도록 한다. 또한 여과과정에서 탈락하여 사용자에게 제공되지 않는 문서 내의 키워드에 대한 가중치도 하향 조정하여 중요도를 낮추거나 궁극적으로 해당 키워드가 제거될 수 있도록 한다. 이 때, 가중치 감소율은 여과 과정에서 선택된 문서에서의 부정적 반응에 의한 감소율보다 더 크게 하였다. 학습과정을 마친 후에는 프로파일 내의 가중치가 0 이하인 키워드를 프로파일 내에서 제거한다. 키워드의 감소는 다음 단계에서 문서를 벡터로 표현하는 과정에서의 계산량의 감소를 기대할 수 있으며, 사용자의 관심의 변화를 반영한 것으로 볼 수 있다.

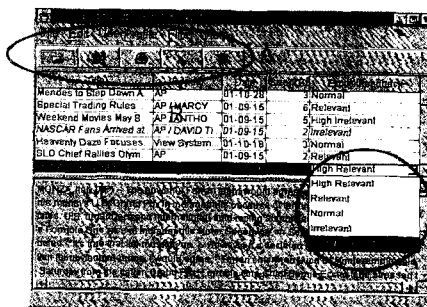
제 4장 실험 및 평가분석

4.1 시스템 구성도

전체 시스템은 4개 부분으로 이루어졌다. 서버에서 뉴스 기사를 가져오는 문서수집 부분, 뉴스 기사를 전처리하여 벡터공간으로 표현하는 문서전처리 부분, 사용자 프로파일과 유사도 비교를 통해 문서의 순위를 계산하는 문서 순위부여 부분, 사용자의 결합된 피드백을 받아 프로파일을 학습하는 프로파일 학습 부분이 유기적으로 상호연관하면서, 하나의 전체적인 정보여과 에이전트를 구성한다.

4.2 사용자 인터페이스

사용자 인터페이스는 Java2(JDK 1.3)로 구현하였으며, 브라우저 형태로 시스템에 접근할 수 있다(<그림 1>). 사용자 인터페이스 화면은 세 개의 프레임으로 구성되어 있다.



<그림 1> 사용자 인터페이스

상위 프레임(A)은 메뉴가 나열되어 있으며, 사용자가 암시적인 피드백 행위를 할 수 있도록 읽기, 저장, 인쇄, 삭제의 메뉴가 있으며, User Profile 메뉴는 사용자의 프로파일 유지와 관련 있는 메뉴로서 메뉴를 열게 되면 사용자가 관심있어 하는 키워드의 리스트를 삽입할 수 있게 해주는 dialog가 펼쳐 보이게 되어 있다. 이 dialog를 통하여 키워드 리스트로부터 키워드들의 삭제 또는 추가등의 편집을 할 수 있도록 하였다. Filter 메뉴는 적재된 각 기사들이 가중치 순서대로 정렬될 수 있도록 하는 메뉴이다

중간 프레임(B)은 사용자에게 뉴스기사 리스트를 보여주는 프레임이며 우측에는 명시적인 피드백을 줄 수 있도록 기사에 대한 5가지 반응이 주어져 있다. 마지막 프레임(C)은 기사를 선택했을 경우 해당 기사의 내용이 콘텐츠 프레임에 나타나도록 되어 있다.

4.3 실험 데이터

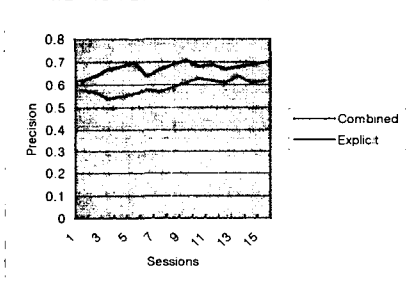
본 논문 실험에 사용되어진 문서들은 USENET 뉴스기사들 중에서 발췌한 것으로 유즈넷 뉴스 그룹 중 clari.web.* 뉴스 기사를 대상으로 하였다. Clari.web.* 뉴스 기사는 ClariNet(<http://www.clari.net>)이 AP, AFP등에서 모은 기사들을 biz(business news), living(entertainment news), local(usa local news), news, sports, tw(science and technology news), usa, world의 8개 카테고리, 200개정도의 세부 카테고리로 분류하여 제공하는 뉴스서비스이다. 실험 데이터는 제목과 본문을 담고 있는 영어 텍스트로 된 문서로써, 2가지 실험을 수행하기 위하여 총 890개의 문서를 사용하였다.

<표 3> 실험 데이터 구성

Usenet News	문서개수
clari.biz.* 계열	382
clari.sports.* 계열	168
clari.living.* 계열	116
clari.local.* 계열	64
clari.tw.* 계열	160

본 논문에서는 실험을 통해 평가하고자 하는 내용은 2가지이다. 정보여과 에이전트가 지녀야 할 성질로서 Specialization 성능과 Adaptation 성능을 확인하기 위한 실험이므로 실험1은 관심영역이 2개 또는 그 이상인 사용자에 대한 실험으로 사용자의 관심영역이 clari.biz.industry.*와 clari.living.music 분야에 관심있는 사용자라 설정하고, 수행하였다. 이 분야와 관련된 키워드를 삽입하고, 사용자가 관심있는 관련 문서에 대해서는 균등한 결합된 피드백을 제공하였다. 실험에 사용한 데이터는 총 300개의 기사이며, 각 기사의 영역은 <표 3>의 영역 모두를 사용하였다

실험1 사용자 관심영역이 복수개인 상황에서의 성능 측정

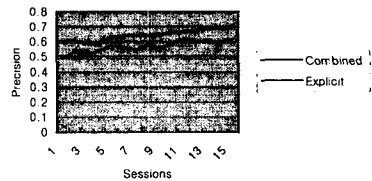


<그림 2> 실험1의 여과성능

15회의 실험결과 평균 Precision은 Combined Feedback의 경우 67.5%, Explicit Feedback의 경우 59.2%의 Precision을 보여 주었다.

실험2는 다수개의 관심영역을 가지며, 학습과정 중 관심영역에 변화를 일으키는 상황에 대한 실험으로 사용자는 초기 clari.biz.industry.*와 clari.living.music에 관심을 보였다가 후에 clari.local.*와 clari.tw.*로 관심이 전이된 사용자이다. 초기에는 clari.biz.industry.*와 clari.living.music에 긍정적인 피드백을 보이다가 새롭게 등장하는 clari.local.*와 clari.tw.*에 긍정적인 피드백을 준다. 실험에 사용한 데이터는 총 300개의 기사가 제공됐다.

실험2 사용자 관심영역의 전이상황에서의 성능 측정



<그림 3> 실험2의 여과성능

15회의 실험결과 평균 Precision은 Combined Feedback의 경우 64.6%, Explicit Feedback의 경우 56.9%의 Precision을 보여 주었다.

제 5장 결론

본 논문에서는 온라인 뉴스 기사를 여과하여 사용자와 관련이 있는 뉴스 기사만을 선별하여 보여주는 정보여과 뉴스 에이전트를 설계, 구현하였다. 이 분야는 다양한 연구가 진행되고 있는 분야로 본 논문에서는 사용자 프로파일 구축에 있어 사용자 피드백 획득방법에 연구 초점을 둬으로써, 기존 단일화된 피드백만을 사용할 때보다 좀 더 나은 정보 여과를 얻으려는데 연구 목적을 두었다. 주로서의 명시적인 피드백과 보조적인 측면으로서의 암시적인 피드백 모두를 결합하여 사용함으로써, 좀 더 나은 사용자의 적합성 피드백을 취합하려 하였다. 이로써 명시적인 피드백의 가장 큰 단점인 사용자에게 부여되는 의식적인 부담감을 최소화하고, 사용자가 실수 혹은 고의로 명시적인 피드백 선택행위를 하지 않더라도, 시스템이 사용자의 암시적 피드백을 감지하는 것을 가능하게 하여 결국은 어떠한 상황 속에서도 사용자의 적합성 피드백을 최대한 얻어낼 수 있었다.

참고 문헌

[1] 이정수 “적합성 피드백을 이용한 정보여과 에이전트” 서강대학교 대학원 석사논문, pp19-30, 1997

[2] 서영우 “강화 학습 기반의 웹정보 여과” 서울대학교 대학원 석사논문, pp21-28, 2000

[3] Dwi H. Widyantoro, “Dynamic Modeling and Learning User Profile in Personalized News Agent”, Texas A&M 대학원 석사논문, pp18-46, 1999

[4] 홍연주, “쇼핑 에이전트를 위한 다차원 사용자 모델링” 서강대학교 대학원 석사논문, pp11-15, 2000