2000년 한국정보처리학회 추계 학술발표논문집 제7권 제2호

지능형 통합 에이전트의 구현

이승원*, 류제**, 유성규**, 한광록**
*호서대학교 컴퓨터공학과 대학원
**호서대학교 벤처전문대학원 컴퓨터응용기술 분야
lhonesty@office.hoseo.ac.kr

The Implementation of Intelligently Integrated Agent

Lee Seung Won*, Ryu Je**, Yu Sung Gyu**, Han Kwang Rok**
*Computer Engineering Hoseo University
** Graduate School of Venture, Hoseo Univ.

요약

정보의 홍수 속에서 사용자가 원하는 정보를 제공하기 위해서는 현재의 Web 상에서의 단순 검색에 의한 정보의 제공이 아닌 뉴스 서버에서의 정보제공과 함께 검색된 정보를 필터링하고 요약하여 더 높은 수준의 정보를 제공하여야 한다.

본 논문에서는 HTTP 와 NNTP를 사용하여 웹과 뉴스 서버상에 존재하는 정보를 검색하고 각 사용자의 Profile을 구축, 이것을 이용하여 피드백, 클러스터링 등의 필터링 기법을 적용, 사용자에게 정보를 제공할 수 있는 지능형 통합 에이전트 시스템의 구현 방법을 소개한다. 제안하는 시스템은 사용자의 관심도를 나타내는 Profile을 이용한 검색된 정보와의 필터링 과정을 통하여 검색된 정보의 유용성과 질을 높일 수 있으며 각 필터링 방법의 모듈화를 통하여 각 학습방법의 개선을 용이하게 한 것이 특징이다.

1. 서론

정보 제공 시스템은 다양한 기술의 발전과 사용자의 새로운 요구에 의해서 급속히 성장하고 있다. 이와 같은 성장은 하드웨어 기술, 통신 기술, 그리고 사용자 인터페이스(User Interface)의 발전으로 인하여 지속될 것이며, 이미 여러 분야 즉, 전자도서관, 원격 교육, 전자상거래 등에서 정보 시스템을 활용하고 있다[1]. 이러한 정보 시스템은 사용자가 원하는 정보를 검색 및 제공하여야 한다. 그러나, 정보를 표현하는 용어의 상이성 및 일반 텍스트를 기반으로 하는 기존의 검색방법 등에서는 많은 문제점을 안고 있는 것이 사실이다. 뿐만 아니라, 사용자가 수행하는 작업에 관한 학습의 부재와 그 활용이 미비함으로 인하여 기존의 정보 시스템은 사용자가 원하는 정보를 효과적으로 제공하기 어렵다. 이러한 기존의 검색 엔진과 같이 사용자가 원하는 정보만을 단순 검색 제공하는 것을 넘어 검색된 정보를 필터링하고 요약하여 더 높은 수준의 정보를 제공하는 것이 필요하다. 이를 위해서는 사용자와 검색된 정보 사이에 매개체가 있어야 한다[1].

본 논문에서는 이러한 매개체 역할을 수행할 수 있는, 웹 정보와 뉴스 서버에서의 정보 검색과, 검색된 정보에 대한 사용자 Profile을 이용하여 필터링을 수행, 사용자가 원하는 정보를 제공할 수 있는 지능형 통합 에이전트 시스템의 구현을 설명한다.

본 논문에서 구현한 시스템은 웹 에이전트(Web Agent)와 뉴스 에이전트(News Agent)가 웹서버와 뉴스 서버상의 정보를 추출하여 구축되어 있는 각 사용자의 프로파일(Profile)을 이용하여 피드백, 클러스터링 등의 방법을 이용하여 필터링을 수행, 기존의 단

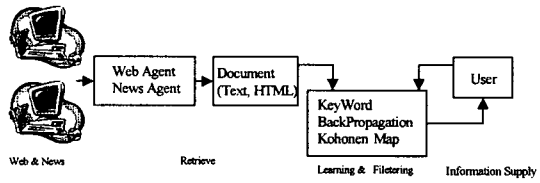
순 정보 제공이 아닌 향상된 수준의 정보를 제공할 수 있게 된다.

2. 관련연구

에이전트 기법을 사용한 분야에서, 현재 활발한 연구가 이루어 지고 있는 분야는 기존의 웹 검색과 더불어 전자상거래, 전자 도서관 등이 있다. 특히 전자상거래에서의 지능형 서비스 제공은 주로 지능형 에이전트의 활용을 통해서 구현되고 있다. Anderson Consulting 의 Bargain Finder[8], 휴리스틱 탐색과 귀납적 학습기법을 적용한 Shopbot 등이 그것이다. 또한 Metacrawler, Excites 등 인공지능 기법을 이용한 정보 분류 기능을 갖는 검색엔진이 개발되어 사용되고 있으며 MIT에서 개발한 Amalthea 는 멀티 에이전트 시스템으로 구성된 검색엔진으로, 웹을 탐색하는 웹 검색엔진, 웹에서 특정 정보를 찾아내는 정보검색 에이전트, 찾아낸 정보를 적절하게 추려내는 정보필터링 에이전트, 그리고 사용자 인터페이스로 구성되어 있으며 특정 사용자의 기호에 맞는 정보를 제공할 수 있는 방향으로 학습 가능한 것이 특징이다[2][7][9]. 그러나 뉴스 서버상의 정보와 더불어 웹 문서에 대하여 사용자 프로파일을 이용한 필터링을 수행하는 시스템은 많지 않은 실정이다. 따라서 본 논문에서는 지능형 통합 에이전트를 이용하여 웹과 뉴스 서버의 정보에 대하여 사용자 프로파일을 이용, 다양한 필터링 기법을 적용해 보았다.

3. 통합 에이전트의 설계

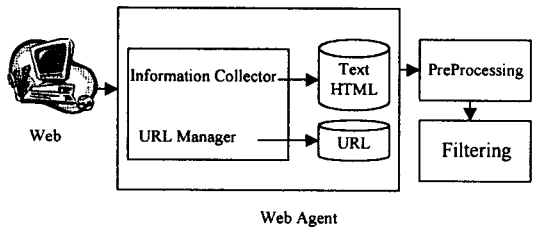
본 논문에서 구현한 시스템의 전체 블럭도는 (그림 1)과 같다. 웹과 뉴스 서버상의 정보들을 웹 에이전트와 뉴스 에이전트를 이용하여 수집하고 각 해당 정보들을 사용자의 관심도와 함께 필터링 과정을 거쳐 사용자가 원하는 정보를 제공하는 것이 본 시스템의 목적이다. 본 시스템은 크게 웹 에이전트와 뉴스 에이전트의 검색 부분, 필터링 부분, 정보제공 부분으로 구분할 수 있다.



(그림 1) 시스템 블럭도

3.1 웹 에이전트

본 논문에서 구현한 웹 에이전트의 주된 기능은 웹 서버상의 정보를 가져와 HTML 태그 제거 등의 전처리 과정을 거쳐 필터링 부분의 입력으로 보내는 것이다. (그림 2)는 본 논문에서 구현된 웹 에이전트의 구조를 나타낸다. 웹 에이전트는 URL Manager, Information Collector의 모듈로 구성되어 있다. Information Collector는 웹 서버에 접속을 하여 웹 문서를 수집하고, URL Manager는 웹 문서에 링크되어진 웹 서버의 주소를 추출하고 방문 및 미방문 URL을 등록한다[3]. 이 웹 에이전트의 출력이 뉴스 에이전트의 출력과 더불어 필터링 부분의 입력이 된다.



(그림 2) 웹 에이전트

3.2 뉴스 에이전트

웹과 함께 본 논문에서 적용한 것은 뉴스 서버상의 정보제공이다. 뉴스 서버상의 정보는 해당 뉴스 서버에 접속하여 NNTP 를 이용하여 뉴스 내용을 가져오며(RFC977,RFC850) 본 시스템에서는 119 포트에 연결하여 자바의 소켓 클래스를 이용하여 정보를 가져오게 된다.

3.3 필터링

사용자가 원하는 정보를 다수의 원시 데이터에서 어떻게 가공하여 사용자에게 있어 정보로서의 가치를 부여하도록 할 것인가는 정보제공 시스템을 구성함에 있어 가장 핵심적인 내용이다. 수많은 정보로부터 사용자의 관심도와 가장 밀접한 관계를 나타내는 문서를 제공하기 위하여 필터링 기법이 적용이 되는데 본 논문에서는 Kohonen-Map 과 BackPorpagation 알고리즘을 적용했다[4].

4. 구현 및 실험

위의 알고리즘을 이용하여 신경망을 구성하기 위한 초기의 데이터는 처음 뉴스그룹에 접속하여 무작위로 20개의 기사 내용을 가져와 사용자 프로파일과 키워드 매칭을 하여 얻은 데이터를 정규화(normalized)하

여 얻는다. 이렇게 정규화한 데이터에 weight와 activation을 적용시켜 각 알고리즘에 해당하는 네트워크를 구성한다. 무작위로 선택된 20개의 뉴스 기사를 가지고 BackPorpagation 알고리즘으로 필터링을 적용했을 때, 각 기사에 실린 키워드의 수와 정규화 결과, 그리고 weight와 activation을 적용시킨 결과는 (표1)과 같다.

(표1) 키워드수와 Desired Value

KeyWork		jdbc	애플릿	Servlet	Jsp	Desired Value
기사 원문에 나타난 키워드 수	기사 1	0	0	0	0	0.0
	기사 2	0	0	0	0	0.0
	기사 3	0	0	1	0	0.25
	...	...	...	...	...	...
	기사 19	0	0	0	0	0.0
	기사 20	0	0	0	1	0.25

위 표에 따른 각 기사의 정규화 결과는 다음과 같다.

Record1: 0.0 0.0 0.0 0.0 0.0 0.0
Record2: 0: 0.0 0.0 0.0 0.0 0.0 0.0
Record3: 0: 0.0 0.0 1.0 0.0 0.33333334 0.2
...
Record19: 0.0 0.0 0.0 0.0 0.0 0.0
Record20: 0.0 0.0 0.0 0.5 0.33333334 0.2

위의 정규화 데이터에 weight와 activation을 적용시킨 결과는 다음과 같다.

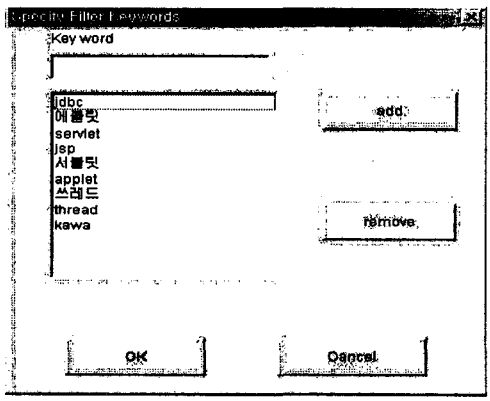
activations= 0.0 0.0 0.0 0.0 0.0 0.6409174 0.6013485 0.47930434 0.6175698 0.5912287 0.67295057 0.63589555 0.63531154 0.6521856 0.6533997 0.4503265
Desired: 0.0 Actual: 0.4503265

activations= 0.0 0.0 0.0 0.0 0.0 0.6409174 0.6013485 0.47930434 0.6175698 0.5912287 0.67295057 0.63589555 0.63531154 0.6521856 0.6533997 0.4503265
Desired: 0.0 Actual: 0.4503265

activations= 0.0 0.0 1.0 0.0 0.33333334 0.7206338 0.75510937 0.72486746 0.722004 0.66464686 0.7026865 0.6911424 0.7184002 0.7983515 0.8075011 0.5033468
Desired: 0.25 Actual: 0.5033468

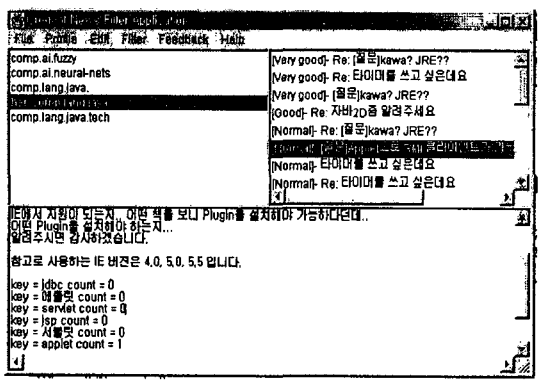
BackPropagation에서 사용자가 작성한 프로파일(관심도의 표명)에 따라 일치된 키워드의 개수로 일차적인 출력값(desired value)을 구한다. 본 논문에서는 일치된 키워드가 없는 기사는 불필요한 기사로 간주하고 0.0 이라는 값을 할당했다. 그리고 점차적으로 키워드가 많이 나올수록 사용자가 원하는 기사라고 간주하고 0.25, 0.50, 0.75, 1.0 의 출력값을 부과한다. 일단 이러한 신경망이 구성되면 학습시킨 weight 값을 가지고 새로 들어온 데이터를 네트워크 망에 삽입하여 원하는 결과치와 비교하여 해당기사를 '아주 나쁨', '나쁨', '보통', '좋음', '아주 좋음' 등으로 구분한다. 각 기사에 대하여 다음의 처리과정을 거쳐 결과를 구성한다.

```
에러값이 일정 범위 안에 들어 오면 만족할만한 결과라고 여긴다.
for(i = firstOut ; i < numUnits ; i++) {
        error[i] = teach[i-firstOut] - activations[i] ;
       // output error 계산
        sumSquaredError += error[i] * error[i] ;
       // squared error 합
        if (Math.abs(error[i]) < tolerance) error[i] = (float)0.0 ;
        // 만족한 결과를 찾음
       deltas[i] = error[i] * activations[i] * (1 - activations[i]);
}
```



(그림 3) 사용자 키워드 입력화면

(그림 3)의 화면은 사용자가 자신의 관심도를 표명할 수 있는 화면이다.

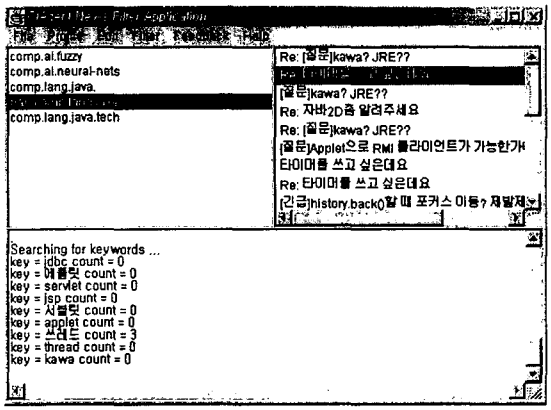


(그림 4) BackPropagation 적용결과 화면

(그림 4)에서 사용자는, 접속한 뉴스그룹에서 가져온

정보를 (그림 3)의 화면에서 사용자가 표명한 관심도에 따라 해당 정보를 [Very Good], [Normal] 등의 필터링 결과와 함께 제공되는 것을 볼 수 있다.

키워드의 출현 빈도에 따른 필터링 결과는 다음 그림과 같다. (그림 5)의 화면에서 사용자는 해당 문서에서 사용자가 표명한 관심도, 즉 키워드의 일치 개수에 따른 결과를 볼 수 있다.



(그림 5) 키워드 출현 빈도에 따른 필터링 결과

(그림3)과 (그림4)에서 보는 것과 같이 사용자의 관심도에 따라 학습되어 있는 각 네트웍의 결과에 따라 필터링된 기사의 가중을 줄 수 있음을 알 수 있다. 이 결과에 따라 자신의 관심도를 표명한 사용자에게 해당 문서를 제공하게 된다[6].

5. 결론

지금까지 본 논문에서 구현한, 웹과 뉴스 서버상의 문서들을 BackPropagation과 키워드 등의 필터링방법을 통하여 사용자의 관심도에 따른 문서 제공 시스템에 대하여 알아보았다.

각 필터링 방법을 통하여 학습된 각각의 네트웍들은 새로운 문서가 들어올 때마다 사용자의 피드백 또는 사용자 행동의 모니터링을 통하여 계속해서 학습해 갈 수 있으며 이러한 결과를 통하여 사용자의 관심도와의 일치성을 얼마나 높일 수 있는가 하는 것과 사용자 프로파일 관리가 향후 과제가 되겠다.

또한 단순히 각 에이전트들이 가져온 문서를 그대로 사용자에게 제시하는 것이 아니라 문서 요약 시스템을 통하여 해당 문서의 내용을 추출, 요약하여 사용자에게 제공하여 사용자로 하여금 문서의 선택을 도울 수 있는 시스템 또한 향후과제로 진행 중에 있다. 이러한 문서 요약 시스템은 키워드의 조작에 따라 해당 문서의 핵심문장을 추출 제공함으로서 사용자가 제공된 모든 문서를 읽지 않아도 원하는 문서를 좀더 빠른 시간에 찾아낼 수 있으므로 본 논문에서 구현한 시스템에 추가된다면 좀더 발전된 정보제공 시스템의 모델이 될 수 있으리라 기대된다[5].

참고문헌

[1] Gio Wiederhold, "Extrapolating Trends for Information Technology" 첨단정보기술연구소 소식지 2호, 2000.

[2] Alexandros Moukas, "Amalthaea: Information Discovery and Filtering using a Multiagent Evolving Ecosystem", The proceedings of the Conference on Practical Applications of Agents and Multiagent Technology, London, April, 1996.

[3] Gun-Woo Nam, Jong-Hee Park, Tai-Yun Kim, "Dynamic Management of URL Based on Object-Oriented Paradigm", IEEE, 1998.

[4] 선복근, "신경망 학습을 이용한 문서자동분류를 위한 학습에 관한 연구 ", 인간과 컴퓨터 상호작용 연구회 회보, 9권 1호, pp925-930, 2000.

[5] 류 재, 선복근, 박보아, 한광록, "단어의 공기관계 그래프를 이용한 문서 요약 시스템의 구현", 정보과학회 춘계학술논문집, 27권 1호, pp348-350, 2000.

[6] Joseph P. Bigus and Jennifer bigus, "Constructing Intelligent Agents with Java", Wiley Computer Publishing, 1998.

[7] 이정은, 임미혁, "전자상거래 상에서 쇼핑 에이전트", http://cybernet.pe.kr/search/robot/shopagent.html

[8] http://bf.cstar.ac.com/bf.

[9] 이경전, "전자상거래 소프트웨어 에이전트", 정보처리학회지, Vol.6 No.1, 1999.