

리서치 에이전트 시스템에서의 지능적 프로파일 구축을 위한 개선된 확산 활성화 네트워크

조영임[†] · 김유신^{**}

요 약

과학기술 연구는 인터넷에서부터 최신의 정보를 필요로 한다. 하지만 인터넷의 방대한 양의 정보를 검색하고 개인의 연구자에 따른 정보를 추출해 내는 것은 복잡하고 반복된 과정이다. 이 논문에서 우리는 이러한 리서치 과정을 돕는 에이전트(PREA)시스템을 제안한다. 이 시스템은 주로 사용자의 관심사항을 시스템을 사용하는 과정에서 관찰하고 지능적으로 프로파일을 구축하게 된다. 그러나 프로파일을 사용자 중심으로 보다 시각적, 실시간으로 구축하기 위해서 이 논문에서는 기존의 확산 활성화 네트워크(Spreading Activation Network)를 개선한 확산 활성화 네트워크(MSAN: Modified Spreading Activation Network)를 이용하고자 한다. 이렇게 함으로써 사용자의 관심사를 의미망으로 구축해서 보다 효율적인 정보검색이 가능해진다. 또한 관련된 여러 실험을 하였는데, 그 결과 일반적인 상업 웹 검색엔진보다 PREA시스템이 맞춤형 정보검색에 보다 효율적임을 알 수 있었다.

Modified Spreading Activation Network for Intelligent Profile Construction in Research Agent System

Young-Im Cho[†] and Yoo-Shin Kim^{**}

ABSTRACT

The research of science and engineering needs the latest information from internet resources. But it is a complex and repeated procedure to search and filter web documents from the huge Internet resources. In this paper, we propose the PREA system, which can organize the research paper databases and search World Wide Web documents that the user is interested in. It observes the usage of the local paper databases and presented web documents and then constructs a profile intelligently. However, to make a profile, we used the modified spreading activation network(MSAN) so that the PREA can search and filter web documents by semantic meaning of user's interest in realtime. The system constructed in multi-agents manner that can cooperate together effectively. The results show the effectiveness of our system to search web documents compared with a commercial search engine.

Key words: Personalizes Research Software Agent, Semantic Network, modified spreading activation network

1. 서 론

현재 연구되고 있는 대부분의 에이전트들은 효과

이 논문은 2001.9~2002.2까지 평택대학교 교내과제 지원으로 수행되었음.

접수일: 2003년 3월 14일, 완료일: 2003년 4월 14일

[†] 정회원, 평택대학교 컴퓨터학과 조교수

^{**} 준회원, 부산대학교 전자공학과 교수

적으로 작동하기 위해서 사용자들의 정보를 필요로 하게 된다[1]. 특히 개인 맞춤형 웹 서치 에이전트 시스템은 사용자 정보요구 사항을 분석하기 위해 프로파일을 만들게 되는데, 이때 사용자의 관심사를 얼마만큼 지능적이면서 자동적으로 프로파일로 구축하느냐 하는 것이 중요한 문제가 된다[2].

현재까지 사용자 성향을 학습하기 위한 알고리즘

과 시스템이 개발되고 있는데, 대표적인 것으로는 WebWatcher[3], Krakatoa Chronicles[4], Fab[5], News Weeder[6] 등이 있다. 이런 시스템은 사용자 관심도와 웹문서들을 표현하기 위해 벡터 스페이스 모델을 주로 사용하고 적합성 피드백(relevance feedback)과 강화 학습(reinforcement learning) 같은 학습 알고리즘을 프로파일을 구축하는데 이용하여 왔다. 그러나 이런 시도들은 일반적으로 정보 검색 분야에서 예전부터 많이 사용되어 왔지만 단어와 단어 사이 혹은 문서와 문서 사이의 의미론적(semantic) 관계를 표현할 수 없다는 단점을 지니고 있다. 따라서 이 논문에서는 사용자의 관심도를 효과적으로 표현하기 위해 네트워크 기반의 프로파일 구축과 학습 알고리즘을 연구하여 적용하고자 한다.

확산 활성화 네트워크(Spreading Activation Network)는 컴퓨터 공학의 인공지능 분야에서 의미망의 처리 구조로 처음에 제안되어 사용되어 왔다. 기본적인 확산 활성화 네트워크는 다양한 시스템에 적합하게 그 형태가 변형되거나 개선되고 있다[7]. Jennings와 Hideyuki는 개인 맞춤형 뉴스 서비스를 위해 프로파일 학습 알고리즘으로 사용자 기반의 신경회로망을 제안했는데[8], 이는 단순한 SAN을 에이전트 시스템에 처음 적용시킨 예가 되었다. 그리고 Fabio와 Puay는 WebSCA 웹검색 시스템으로 Constrained Spreading Activation을 사용하기도 했다[9].

그러나 확산 활성화 네트워크는 네트워크 구조를 형성하고는 있으나 사용자 중심의 시각화, 실시간처리, 데이터베이스 수정 등 사용자 중심의 맞춤형 기능 제공이 부족하다.

따라서 이 논문에서 기존의 확산 활성화 네트워크를 확장한 개선된 확산 활성화 네트워크(MSAN: Modified Spreading Activation Network)을 이용하여, 논문 서지정보를 관리하고 관련된 웹문서를 자동적으로 찾아주는 Personalized Research Agent (PREA) 시스템을 구축하고자 한다. 이 시스템의 특징은 개인 맞춤형 정보를 실시간으로 구축이 가능하며 다음과 같은 장점을 갖는다.

첫째, 사용자의 프로그램 사용 관찰이 가능하다. 즉, 기존의 확산 활성화 네트워크에서 부족했던 실시간 사용자 서비스를 위해 PREA는 사용자가 응용 프로그램을 사용하는 것을 관찰하고 그것을 프로파일

구축에 이용한다.

둘째, SAN의 수학적 공식화가 가능하다. 즉, 웹 검색에 맞도록 SAN의 각 노드와 그 사이의 가중치를 공식화하여 이것을 가능하도록 구현하였다.

셋째, SAN의 시각화가 가능하다. 즉, 기존의 확산 활성화 네트워크가 시각적이지 못하다면, PREA는 네트워크를 사용자들에게 명시적으로 보여줌으로써 직접적으로 네트워크를 수정 가능하게 만들어준다.

2. PREA 시스템의 구조

그림 1은 PREA 시스템의 구조를 보여준다. PREA는 4개의 에이전트로 모듈과 MSAN 그리고 MSAN을 관리하는 프로파일 관리자 그리고 로컬 데이터베이스 등으로 구성되어 있다. 각각의 에이전트들은 서로 통신을 하면서 사용자 성향을 파악하고 MSAN을 이용해 웹검색 및 필터링을 하게 된다. 각각의 에이전트에 대한 설명은 다음과 같다.

○ 사용자 행동 에이전트(User Behavior Agent): 응용 프로그램을 사용하는 사용자의 행동, 특히 1) 사용자가 웹 서치 에이전트를 통해 제시된 웹문서를 선택하거나 거부할 때 2) 사용자가 MSAN을 시각화 에이전트를 통해 수정할 때 3) 사용자가 내부의 논문 서지 정보 등을 검색할 때 등을 관찰한다. 그리고 그 정보를 프로파일 매니저에게 전송한다.

○ 시각화 에이전트(Visualization Agent): 사용자가 요구를 할 때 MSAN을 노드와 그 사이의 가중치로 시각화해서 보여준다.

○ 웹서치 에이전트(Web Search Agent): 인터페이스 에이전트가 요구를 해오면 MSAN으로부터 질의어를 자동적으로 생성하고 검색 엔진을 통해 웹문서 검색을 한 뒤 MSAN을 통해 검색된 문서에 각각 사용자의 관심도에 부합하는 순서로 순위를 정한다(3장 참조). 그리고 그 결과를 인터페이스 에이전트에 전달한다.

○ 인터페이스 에이전트(Interface Agent): 사용자와 다른 에이전트들 간의 정보 요구와 응답을 담당하며 내부 서지 정보 검색시스템과 관리를 담당한다.

○ 프로파일 매니저(Profile Manager): 사용자 행동 에이전트가 요구를 할 때 MSAN을 수정하거나 갱신시킨다.

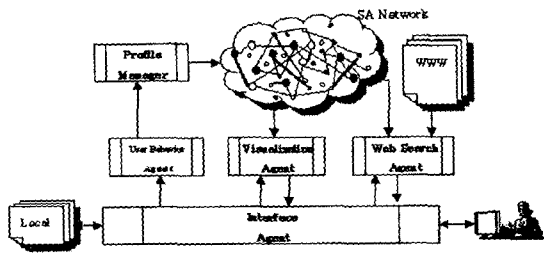


그림 1. PREA의 구조

3. 개선된 확산 활성화 네트워크: MSAN

확산 활성화 네트워크(Spreading Activation Network)는 정보 검색 분야에서 사용되어 연상 처리 패러다임(associative processing paradigm)이다[8]. 네트워크는 그림 2에서와 같이 링크로 연결된 노드들로 구성되어 있다. 노드 i 의 에너지 W_i 는 전체 네트워크 상에서 단어 i 가 가지는 중요도를 나타내고 W_{ij} 는 단어 i 와 j 사이의 상관 가중치(correlation weight)로서 두 단어 사이의 관련도를 나타내는 것이다.

노드 에너지와 상관 가중치는 내부 서지 데이터베이스의 사용결과와 사용자가 선택하거나 거부한 웹 문서들에 의해서 갱신된다. 네트워크는 온라인으로 갱신되는 것이 아니고 오프라인으로 한 세션마다 갱신되므로 네트워크 구축에 걸리는 계산 시간은 그렇게 크지 않다. 네트워크는 변화하는 사용자의 성향을 반영해야 하므로 시간이 지남에 따라 노드의 수, 에너지와 상관 가중치 등이 변하거나 새로운 노드가 생기기도 한다.

그러나 확산 활성화 네트워크를 구축하고 갱신하는 방법이 명확하지 않고 실시간으로 이루어지지 않기 때문에 사용자 중심의 프로파일 작성에 정확도가 떨어지는 점을 개선하기 위해, 이 논문에서 MSAN

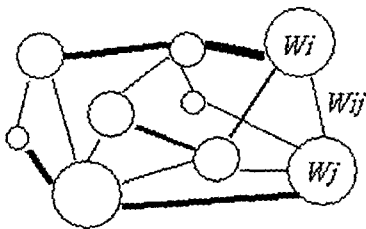


그림 2. 확산 활성화 네트워크(Spreading Activation Network)

을 구축하고 갱신하는 두 가지 방법 - 내부 서지 정보에서 그리고 검색된 웹문서를 통해 - 을 수학적으로 규명하고, 웹문서를 랭킹하는 방법을 확산 활성화 네트워크의 확산 과정을 통해 구축할 예정이며, 이러한 과정 및 구조를 이 논문에서는 개선된 확산 활성화 네트워크(MSAN: Modified SAN)라고 명명하기로 한다.

3.1 내부 서지 정보에서의 학습

$$W_{iNEW} = W_{iOLD} + t f_i \cdot \alpha^l \cdot F_k \cdot A_i \quad (1)$$

W_{iNEW} 는 노드 i 의 갱신된 에너지이고, W_{iOLD} 는 노드 i 의 갱신 전 에너지이다. 노드의 갱신된 에너지는 이전 에너지에서 특정 추출 리스트에서의 단어 i 의 빈도(term frequency) $t f_i$, 지역 논문서지 정보 상수 α^l , 필드상수 F_k (k : 제목, 저자, 초록) 과 데이터 베이스 접근 상수 A_i (i : 검색, 인쇄, 선택 등)에 의해 업데이트 되어진다.

3.2 웹문서의 학습

$$W_{iNEW} = W_{iOLD} + t f_i \cdot \alpha^s \cdot f \cdot H_k \quad (2)$$

식 2 번은 1번과 비슷하지만 몇 가지 다른 변수들이 있다. α^s 는 웹문서 상수로써 위의 내부 서지 데이터베이스의 단어와 그 에너지를 차이 나도록 하기 위해 쓰인다. f 는 퍼드백 상수로써 사용자가 선택을 한 웹문서의 단어이면 +1, 그렇지 않으면 -1로 할당된다. H_k 는 하이퍼 텍스트의 태그 상수로써 볼드체나 이탤릭체 그리고 링크 부분의 태그가 붙여진 단어에 좀 더 높은 에너지를 준다.

3.3 단어 사이의 가중치 학습

$$W_{ijNEW} = W_{ijOLD} + \alpha C_{ij} \quad (3)$$

단어와 단어 사이의 가중치는 두 단어가 하나의 특징추출 리스트에 자주 같이 나오는 빈도수 C_{ij} 와 그 상수 α 로 갱신 된다. 두 노드 사이의 가중치가 클수록 두 단어는 상관도가 높아지게 되고 나중에 있을 MSAN의 확산을 하는데 있어서 기준이 된다. 그리고 노드의 에너지나 연결된 상관 가중치는 계속 커지는 일이 없도록 일정한 양을 감소하게 만들고 최대 에너지와 최소에너지를 정해 주어야 복잡성을

감소시킬 수 있다.

3.4 MSAN의 확산 과정

위의 과정을 통해서 MSAN이 이미 구축되었고 웹 서치 에이전트를 통해 일반적인 검색엔진으로 웹문서를 찾은 뒤의 그것들의 특징추출 리스트(웹문서의 단어들을)를 만들었다고 하면, 검색된 웹문서를 MSAN의 확산과정을 통해서 사용자의 관심도에 맞게 순위를 정할 수가 있다.

확산 과정 기술은 그림 3에서 보는 것과 같이 MSAN이 평형상태가 되어서 더 이상 발화하는 노드가 없을 때까지 반복적으로 서로 다른 가중치로 연결된 각 노드의 발화가 파동을 이루면서 퍼지게 만드는 과정이다. 확산 과정은 먼저 검색된 웹문서의 특징추출 리스트 안에 들어 있는 단어로부터 시작한다. 일단 그 리스트 안의 단어들이 MSAN의 노드들에 있으면 발화하고 발화된 노드와 연결된 MSAN의 각 노드들에 대해서 그림 3과 같은 조건으로 발화할 수 있는지 분석해 본 뒤에 발화과정이 확산된다. 몇 번의 반복과정을 거쳐 평형상태에 이르면 발화된 노드의 에너지 값을 합해서 순위 값을 매기면 가장 큰 값을 갖는 웹문서가 사용자의 관심도와 가장 일치하게 되는 것이다. 이러한 비선형적인 문서 랭킹 과정은 사용자 관심도에 관한 문서의 탐색을 보다 깊게 하는 효과를 가진다.



$$\text{Firing condition : } \sum W_i W_j > \text{threshold}, \text{ Ranking score: } \sum_{\text{fired node}} W_i$$

그림 3. MSAN의 확산 과정

4. PREA의 구현

그림 4는 PREA 시스템의 초기 화면을 보여준다. 초기 화면은 미리 입력되어 있는 내부 서지 정보를 검색하고 관리할 수 있는 기능을 보여준다. 사용자가 초기에 이 정보를 검색하는 과정은 사용자 행동 에이전트에 의해 기록이 되며 차후 MSAN의 갱신에 이용되게 된다.

초기 화면의 기능을 자세히 살펴보면, 논문 서지 정보를 입력, 편집, 출력 및 검색기능을 가지고 있고 단순한 단어 검색의 결과는 그림 4의 오른쪽 창을

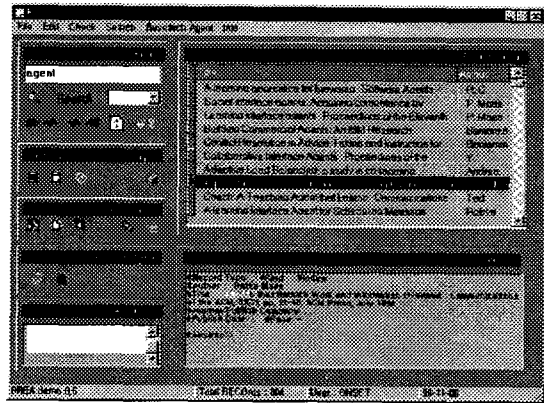


그림 4. PREA의 초기화면

통해 나타내어진다. 그림에서는 'agent'라는 단어로 논문서지 검색을 해 본 것이다. 오른쪽 아래의 정보는 선택한 논문서지 정보의 보다 상세한 내용을 보여주는 창이다.

4.1 프로파일 관리자와 MSAN의 시각화

초기 화면에서 사용자의 논문서지 정보사용의 결과는 MSAN을 초기에 구축하는데 주로 유용하게 쓰인다. 처음부터 웹 검색의 결과만을 가지고 네트워크를 구성하는 것 보다는 훨씬 정확하고 빠른 MSAN 구축이 가능하다. 그림 5는 초기에 구축된 MSAN을 보여주는 화면이다. 초기에는 프로파일이 사용자의 논문서지 정보사용을 기반으로 하지만 나중에는 웹 검색 결과도 함께 3장에서 살펴본 프로파일 갱신 과정에 따라 구축이 된다. 왼쪽의 리스트는 노드의 에너지가 높은 순서대로의 단어들이 나열되어 있고 오른쪽에는 노드와 연결된 다른 노드 그리고 각 노드들

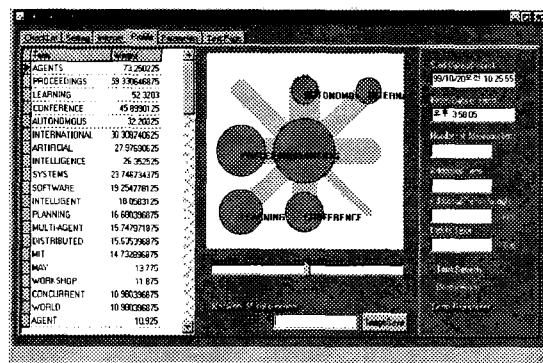


그림 5. 프로파일 관리자

의 에너지와 연결 가중치의 크기까지를 시각적으로 보여주고 있다.

원의 크기는 상대적인 노드 에너지의 크기를 나타내고 연결된 선의 굵기는 노드 사이의 관련도를 나타내는 가중치의 상대적인 크기를 나타내어 사용자가 시각적으로 MSAN을 탐색 가능하게 만들어 준다. 한편 사용자는 프로파일 관리자에서 MSAN의 노드들을 직접 추가 삭제 편집이 가능하다. 하지만 직접적으로 수정을 할 때에는 그것이 전체적인 시스템 성능에 영향을 주므로 신중을 기해야 한다.

4.2 웹문서 랭킹과 브라우징

MSAN이 구축되거나 갱신된 뒤에, PREA는 사용자가 관심 있어 하는 웹문서를 검색하기 위해서 질의어를 생성하고 AltaVista를 통해 웹문서를 검색한다. 그림 6은 이런 과정을 보여주는 화면이다. 질의어 생성은 MSAN의 노드 중 에너지가 높은 것을 순서대로 선택해서 자동 생성한다. 질의어 개수는 사용자가 변경할 수 있으며 그에 따라 검색 결과의 precision/recall이 달라지게 된다. MSAN에서 질의어를 생성 후 그 질의어를 AltaVista를 통해 검색을 하면 많은 문서들이 검색되는데 PREA에서는 그 문서들을 MSAN을 이용해 3장에서 설명한 방법으로 랭킹을 하게 된다.

웹 검색과 랭킹의 작업 시간은 각 웹사이트를 일일이 방문해서 특징리스트를 만들어야 하기 때문에 랭킹될 웹문서의 개수와 네트워크의 상황에 따라 달라지게 된다. 실험에서는 랭킹될 최대의 웹문서를

30개로 제한하였고 웹사이트 방문시간의 데드라인을 설정해서 접속되지 않는 웹사이트의 방문에 시간을 낭비하지 않도록 하였다.

그림 6은 MSAN을 이용해 웹문서가 검색된 결과를 보여준다. 웹문서의 URL과 제목, 웹사이트의 정보와 최종 수정일 등의 정보가 가장 MSAN 랭킹과정에서 높은 점수를 얻은 순서대로 정렬되어 사용자에게 제공된다. 사용자가 URL을 클릭하면 내부 인터넷 브라우저를 통해 해당 웹사이트로 접속을 하게 된다.

PREA는 Microsoft 사의 Internet Explore 6.0을 ActiveX로 내장한 브라우저를 가지고 있다. 사용자가 랭킹된 웹사이트를 방문하는 것을 관찰하고 링크를 통해 사용자가 방문하는 사이트 혹은 유용하지 않다고 선택하는 사이트의 정보를 저장하고 MSAN을 갱신할 때 이 정보를 이용한다. 웹 브라우징 과정이 끝나면 사용자는 프로파일 업데이트를 할 수 있다.

이 과정이 하나의 세션이 되며 초기에 내부 논문 정보 서지 데이터의 사용이 MSAN을 구축을 빨리 하는데 많은 도움이 되고 시간이 지나고 부터는 웹문서의 정보도 MSAN구축에 함께 이용된다.

5. 실험결과

사용자가 PREA의 내부 논문 서지정보 탐색하는 것과 웹에서 검색되어 필터링 된 웹문서에 따라 MSAN이 변화하기 때문에 시스템의 성능을 측정하기 위해서는 실제 사용자에게 의한 실험이 필요하다. 실제 사용들에게 PREA시스템을 사용하게 하고 먼저 MSAN이 사용자의 관심사에 맞게 구축이 되었는지 살펴보고 실제로 웹검색을 통해 웹문서를 랭킹해서 기존의 검색 엔진과의 성능비교를 해 보았다.

5.1 MSAN 구축

MSAN 구축을 실제적으로 하기 위해서 10 세션 동안 시스템을 실제 사용자에게 사용하도록 하였다. 하나의 세션은 웹문서를 검색하고 랭킹하며 네트워크 갱신이 되는 전 과정을 말한다. MSAN의 빠른 구축을 위해서 1, 2, 6, 7번째 세션은 내부 논문 서지 정보를 사용하게 해서 MSAN을 갱신하도록 하고 나머지는 검색된 웹문서를 가지고 갱신하였다. 사용자

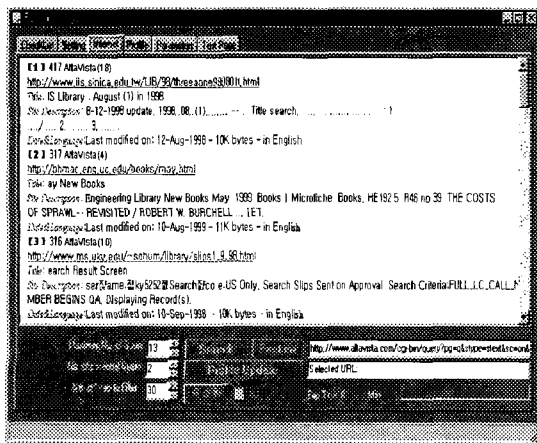


그림 6. 웹문서 랭킹 결과

의 관심은 'software agent'라는 넓은 범위에 있었고 자신도 정확히 'software agent's의 세부적인 분야에 대한 정보와 관심은 초기에 갖고 있지 않았다. 이 논문에서는 실험에 앞서 *Collection of Computer Science Bibliography*에서 1,200개의 'software agent' 관련 논문 서지 정보를 가져와 내부 데이터베이스에 저장하고 사용자가 이를 이용하도록 만들었다. 첫 번째 세션에서 사용자는 먼저 저장된 데이터베이스에 몇 차례 키워드로 논문을 검색해 보게 되고 이것을 바탕으로 초기의 MSAN이 구축된다. 다음부터는 구축된 MSAN을 기반으로 웹 검색과 논문정보 검색을 번갈아 가면서 실시한다.

앞서 살펴본 그림 5는 이 MSAN구축의 실험 결과를 보여준다. 원편은 노드의 에너지가 높은 순으로 'agents', 'proceedings', 'learning', 'conference'로 보여지고 오른쪽 그림은 시각화 에이전트를 통해 선택된 노드들의 연결된 노드와 그 상관도, 그리고 자체 노드의 에너지등도 쉽게 눈으로 확인해 볼 수 있다. 그림 7은 몇몇 노드들 'intelligence', 'distribute', 'agent', 'reinforcement', 'architecture', 'software' 등의 예를 보여주고 있다. 첫 번째 그림을 보면 'agent'라는 노드의 에너지가 다른 것보다 꽤 크고 'intelligence'와 'artificial'의 상관가중치가 크게 나와 있다. 이는 앞서 MSAN의 구축과정에서 설명했듯이 'intelligence'와 'artificial'이 함께 나오는 특정리스트가 많이 있었다는 것이고 이는 나중에 있을 웹문서 랭킹에도 'intelligence'가 발화하면 'artificial'도 같이 발화할 수 있는 가능성이 높다는 것을 보여준다.

MSAN 구축 자체로서의 평가는 수학적으로 내리기 힘들다. 단지 나중에 평가해 볼 웹문서 랭킹에 있어서 기준이 되는 것이고 PREA에서는 사용자가 높

은 에너지의 노드들을 중심으로 확인해 볼 수 있고 직접 수정할 수도 있게 된다.

5.2 웹문서 검색과 랭킹

웹문서 검색과 랭킹을 위해 이 논문에서 앞서서 구축한 MSAN을 이용해 실험을 계속 하였다. 웹문서를 검색하기 위해 이 논문에서는 AltaVista검색엔진을 이용하였다. 먼저 질의어 13개를 MSAN의 가장 높은 에너지를 가진 노드의 단어를 추출해 만들고 검색을 하였고 검색엔진에서 검색된 총 웹문서는 335개였다. 우리는 랭크될 웹문서를 30개로 제한하였으므로 PREA는 AltaVista의 검색 상위 순위 30개를 방문하여 특정 리스트를 만들고 3.4절에서 설명한 방법으로 MSAN확산과정을 통해 각 웹문서를 랭킹하고 다시 순위를 만들어 사용자에게 제시했다.

표 1은 PREA에서 랭킹된 1위부터 20위까지의 순위를 보여주고 있다. 두 번째 행은 외부 검색엔진에서의 순위를 보여주는 것이며 이 표에 나타난바와 같이 PREA와는 다른 순위를 가진다. 그것은 외부 검색엔진은 단지 제시된 질의어가 문서에 들어 있는가를 확률적으로 따진 순위인 반면에, PREA는 미리 구축된 사용자의 관심도가 저장된 MSAN을 통해 랭킹을 하기 때문이다.

예를 들어, AltaVista의 검색랭킹 1위는 PREA에서는 18위에 지나지 않았다. 그리고 PREA의 검색랭킹 1위는 AltaVista의 검색 랭킹 21위에 머물고 있다. 각각의 웹문서를 방문해 분석해 보면 쉽게 그 이유를 알 수 있다. PREA 검색 링크 1위 사이트의 내용은 Twaian Shinca University 'artificial intelligence'에 관한 논문과 학술회의의 목록이다. 문서 안에 MSAN의 많은 노드들을 활성화 할 수 있는 단

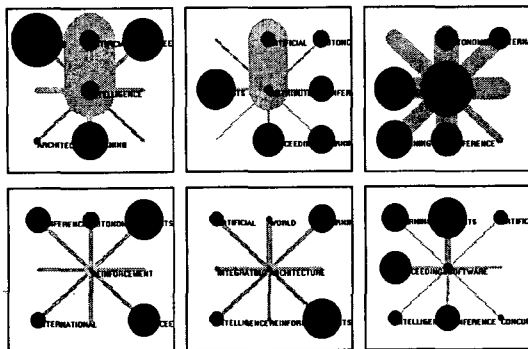


그림 7. 몇몇 노드에 대한 MSAN 시각화

표 1. 랭킹 결과

PREA	AltaVista	Score	URL
1	18	417.11	http://www.iis.sinica.edu.tw/LIB/98/threasaone900011.html
2	4	317.03	http://bhmc.eng.uc.edu/books/may.html
3	10	316.23	http://www.ms.uky.edu/~sohum/library/slips1_9_98.html
4	17	259.95	http://regoc.srce.hr/it/97proc.html
5	12	233.06	http://www.library.cshayward.edu/acq/books/csi.htm
6	6	134.50	http://beta.bids.ac.uk/journalists/rstjournals.txt
7	21	126.92	http://www.library.ucg.ie/floors/fcgisiss.txt
8	5	122.87	http://www.daimi.sau.dk/~brian/EDA/26text
9	2	105.03	http://www.dbd.puc-rio.br/tri_pen11.html
10	22	78.04	http://bioneer.kaist.ac.kr/~yshangj/esci_list.txt
11	14	74.72	http://www.bids.ac.uk/journalists/scjournals.txt
12	15	74.72	http://www.eisss.it/uno/journal.html
13	20	74.72	http://156.17.85.29/8st94.htm
14	30	74.72	http://www.library.kaist.ac.kr/dlibrary/EXPANDED.txt
15	28	68.07	http://www.ifs.hi.fi/ifs/bibliotek/c/cc-i.html
16	29	68.07	http://library.kimm.re.kr/SC/acc-i.html
17	23	67.59	http://cad.kaist.ac.kr/info/SCI_List/A.htm
18	25	62.65	http://www.nova.edu/cwis/hpl/library/bk0698.html
19	3	49.04	http://www.state.nj.us/ESession/WWWBILLL.HTM
20	19	41.89	http://www.ml.uscme.poznan.pl/journals.html
21	1	17.47	http://204.17.96.10/WWWBILLL.HTM

어들이 많이 있었기 때문에 높은 점수를 얻어 랭킹 1위가 되었다. 그리고 AltaVista의 랭킹 1위 사이트는 단순한 키워드 검색 방식의 웹검색 랭킹의 검색 결과로써 비록 'agent'와 'proceeding' 등의 검색어가 들어가지만 컴퓨터 과학 분야의 'agent'와는 전혀 다른 State of Rhode island의 여행 대리인 정보 등이 담겨 있는 웹사이트이다.

랭킹의 순위에 변화를 주는 많은 파라미터들이 있다. 그 중에서 발화 임계치는 가장 중요한 파라미터이다. 낮은 발화 임계치는 특정 리스트가 네트워크에 제시 되었을 때 쉽게 연관 노드들이 활성화 할 수 있으므로 활동적인 네트워크 활성화를 가능하게 해서 재호출(recall)을 높이지만 정확도(precision)를 낮게 만드는 경향이 있고 높은 발화 임계치는 그 반대의 결과를 초래한다.

5.3 실제 다수의 사용자 실험

위의 MSAN 구축과 웹문서 랭킹 실험은 한 명의 사용자를 대상으로 한 실험이고 이번에는 마지막으로 다수의 다른 관심사를 가지고 있는 사용자들을 대상으로 실제로 실험을 해 보았다. 실험은 2주간 10번의 세션을 가지고 10명의 사용자들을 대상으로 실시하였고 컴퓨터 과학 중 다른 관심사를 가지는 사용자들을 대상으로 하였다. 각각의 사용자들을 위해서 5.1의 실험에서처럼 미리 서지 정보 데이터베이스를 관심 범위에서 저장해 놓았다. 사용자들은 뚜렷이 무엇을 찾아야 한다는 요구보다는 자신의 관심 있어 하고 연구하고 싶어 하는 분야를 가지면서 웹문서에서 흥미로운 새로운 정보를 찾기를 바란다는 조건으로 실험에 응하였다. PREA시스템의 효율성을 측정하기 위해서 다음과 같은 두 가지의 평가 기준과 10명에 대한 평균은 다음과 같다.

○ 상위 10 위의 랭킹 웹문서 중에서 실제로 사용자가 관심있어 하는 웹문서의 개수

(랭킹 시험): 평균 9.3 (Altavista - 5.2)

○ 랭킹된 30개의 모든 웹문서 중에서 실제로 사용자가 관심있어 하는 웹문서의 개수 (네트워크 구축 및 질의어 생성 시험): 평균 12.5

첫 번째 기준은 PREA가 얼마만큼 사용자가 흥미 있어 하는 웹사이트들을 상위 랭킹에 순위를 올려놓을 수 있는가를 보여주는 것이고, 두 번째 기준은

PREA가 MSAN을 통해 얼마만큼 질의어를 잘 생성 했는가에 대한 기준이다. 첫 번째 기준에 대한 실험 결과는 평균 9.3으로 상위 10위 안의 거의 모든 문서의 개수가 AltaVista 검색 상위 10위 안의 문서보다도 훨씬 많다는 것을 보여준다.

또한 그림 8을 10명 중의 2명의 실험결과를 각 세션마다 보여주는 그래프이다. 보는 바와 같이 PREA는 웹문서를 MSAN을 통해 보다 효과적으로 랭킹시킨다는 것을 보여준다. 그림 8에서 특이한 사항은 사용자 2번에서 PREA 상위 10개 랭킹 중 관심있어 하는 문서의 개수가 감소하는 경향을 보이다가 다시 증가하는 것을 볼 수 있는데 이것은 사용자의 관심도가 세션동안 변화한 것으로 나타났다. 시스템이 사용자의 관심도의 변화에 따라가려면 일정 시간이 걸려야 함을 보여주기도 한다.

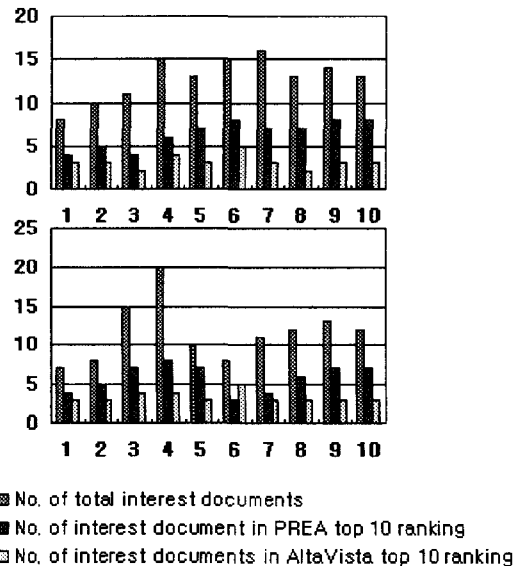


그림 8. 실험결과 중 두 명의 사용자 결과 비교 (상: 사용자 1, 하: 사용자 2)

6. 결론 및 향후 과제

이 논문에서는 MSAN을 기반으로 한 웹문서 필터링을 하는 에이전트인 PREA를 제안하였다. PREA는 각종 논문서지 자료를 관리하는 것 뿐만 아니라 사용자가 관심 있어 하는 웹문서를 찾아 사용자에게 제안하는 기능을 가지고 있다. 우리는 네트워크 기반의 개인 맞춤형 서비스를 효율적으로 하기 위해 사용

자가 프로그램을 사용하는 과정을 관찰할 수 있도록 의미망을 구축하였고 MSAN을 통해 효율적인 웹문서 랭킹을 할 수 있게 만들어서 실제 실험한 결과 사용자의 관심사를 기존의 검색엔진보다 효율적으로 잘 나타내어줌을 실험 결과로부터 알 수 있었다.

MSAN을 이용한 웹정보 처리는 아직 많이 연구되고 있지 않은 분야이고 시스템을 좀더 발전시킨다면 많은 가능성을 가지고 있다. 또한 사용자의 응용프로그램 사용을 관찰해서 그것을 개인 맞춤형 서비스에 이용하는 방법을 적용하기도 하는데 이에 관한 연구는 아직은 시작단계라고 할 수 있다. 이 연구는 계산량의 문제, 휴리스틱 파라미터의 문제 등을 성공적으로 해결해야 하는데, 만약 이것이 가능하다면 본 논문에서 제안한 이 연구는 아주 유용한 개인 맞춤형 웹 검색 에이전트가 되리라고 생각한다. 앞으로 이를 위해 보다 광대한 사용자와 많은 시간을 통해 실험을 해서 MSAN의 변화과정과 기능을 좀 더 명확히 구현하고 다른 방식과 비교해야 할 것이다.

참 고 문 헌

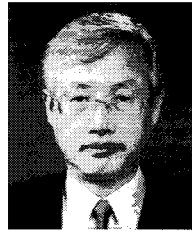
- [1] M.Wooldridge, N.R.Jennings, "Intelligent agents: theory and practice", *The Knowledge Engineering Review*, Vol.40, No.2, pp. 115-152, 1995.
- [2] H.S.Nwana, D.T.Ndumu, "A Perspective on Software Agents Research", *BT Journal*, London, BT Journal, 1998.
- [3] T.Joachims, T.Mitchell, D.Freitag, and R.Armstrong, "WebWatcher: machine learning and hypertext", Fachgruppentreffen Maschinelles Lernen, Dortmund, Germany, August, 1995.
- [4] T.Kamba, K.Bharat, and M.Albers, "An Interactive, Personalized Newspaper on WWW", *In Proceeding of Multimedia Computing and Networking*, 1997, MSAN Jose, California, February 10-11, 1997.
- [5] M.Balabanovic and Y.Shoham. "Fab: Content-Based, Collaborative Recommendation", *Communications of the ACM*, Vol. 40, No.3: pp. 66-72, 1997.
- [6] K.Lang. "Newsweeder: Learning to filter news", *In Proceedings of the 12th International Conference on Machine Learning*, pp. 331-339, 1995.
- [7] S.Preece, "A spreading activation model for Information Retrieval", Ph D thesis, University of Illinois, Urbana-Champaign, USA, 1981.
- [8] Jennings, Higuchi "A Personal News Service Based on a User Model Neural Network", *IEICE transaction on Information and System*, Vol. E75-D No.2, March, 1992.
- [9] F.Crestani, P.L.Lee, "WebSCSA : Web Search by Constrained Spreading Activation", 2001.



조 영 임

1988년 고려대학교 컴퓨터학과 졸업
1990년 고려대학교 컴퓨터학과 석사(인공지능전공)
1994년 고려대학교 컴퓨터학과 박사(인공지능전공)
1996년~현재 평택대학교 컴퓨터

과학과 조교수
1995년~1996년 삼성전자 멀티미디어 연구소 선임연구원
1999년~2000년 University of Massachusetts, at Amherst, Dept. of Computer Science, Post-doc
2003년~현재 (사)한국 퍼지 및 지능시스템 학회 홍보이사 겸 편집위원



김 유 신

1970년 3월~1974년 2월 서울대학교 전자공학 학사
1978년 9월~1980년 6월 U.C. Buckley 전자공학 석사
1980년 9월~1983년 6월 Stanford University 전자재료 박사과정 수학

1992년~1995년 1월 Cornell University Ph. D
현재 부산대 전자공학과 교수, 과학철학회 부회장, 한국통신학회 정보사회 연구회장
관심분야: 인공지능, 신호처리, 과학철학, 뉴로 퍼지

교신저자

조 영 임 450-701 경기도 평택시 용이동 111번지