

사용자 모델 신경회로망을 기반으로 한 사용자 중심의 리서치 에이전트 시스템

송종길*, 김유신, 조영임 jkilsong@hyowon.pusan.ac.kr kimys@hyowon.pusan.ac.kr yicho@ptuniv.ac.kr
부산대학교 전자공학과, 평택대학교 전산학과

Personalized Research Agent System Based on User Model Neural Network

Jong-Gil Song*, Yoo-Shin Kim, Young-Im Cho

Dept. of Electronic Engineering, Pusan National Univ. Dept. of Computer Science, Pyung-Taek Univ.

요약

본 논문에서는 사용자가 자신이 연구하고 있는 분야에 관련된 웹 문서를 스스로 찾아서 보여 주는 PReA 시스템을 구현한다. 사용자의 성향을 파악 하기 위해서 미리 작성된 서지 정보데이터를 사용자가 사용하는 것을 관찰하여 사용자 모델 신경회로망을 구축한다. 사용자 모델 신경회로망은 단어의 부하와 단어 사이의 부하로 구성되어 있어서 사용자의 정보요구의 의미를 나타낼 수가 있다. 사용자 모델 신경회로망을 기반으로 질의어를 생성하고 웹문서를 검색하며 검색된 문서에 대해 순위를 정한다. 순위가 정해진 문서중 사용자가 선택한 문서와 선택하지 않은 문서는 각각 사용자 모델 신경회로망을 학습하는 데 쓰이게 되며 오랜 시간 동안 사용함에 따라 회로망은 사용자의 성향에 적용하게 되어 보다 정확한 검색을 수행하게 된다.

1. 개요

인터넷의 발전에 따라 방대한 정보 속에서 개개인의 사용자가 필요로 하는 정보를 대신에서 검색해주거나 필터링 해주는 에이전트 기술이 부각되어 발전하고 있다.[2] 이런 에이전트 기술의 핵심은 여러 조건에 따라 변하는 사용자의 성향을 얼마만큼 지능적으로 학습할 수 있는가에 달려 있다. 사용자의 성향을 컴퓨터가 이해 할 수 있는 구조로 작성한 프로파일을 학습하는 방법에는 확률적으로 keyword를 분석하는 방법과 사회적 필터링 방법, 그리고 인공지능의 기법을 사용한 것들이 있다.[2,3] 확률적인 방법과 사회적 필터링 방법은 단순하고 이해하기 쉬운 방법인 반면에 문서 내의 문맥과 내용을 파악하기 힘들고 계산량이 많다는 단점을 지니고 있다.

본 논문에서는 향상된 사용자 모델 신경회로망을 기반으로 하여 사용자 프로파일을 구성하고 학습시켜서 웹문서를 검색하고 필터링하는 방법을 제시한다. PReA (Personalized Research Agent) 는 본 논문의 내용을 구현한 시스템으로써 내부에 입력된 논문들의 서지 정보의 사용내역을 관찰하여 사용자가 관심 있어 하는 연구 분야의 웹 문서를 검색하고 필터링한다. 그래서 사용자가 관심 있어하는 연구분야의 웹 문서를 검색하기 위해 많은 시간과 노력이 드는 것을 절약할 수

있다. 웹에서 검색된 문서들은 사용자 모델 신경회로망을 거쳐서 순위가 결정되고 그 중 사용자에게 의해서 선택된 문서들은 신경회로망을 학습하는 데 쓰이게 된다. 그리고 시간이 지남에 따라 사용자 모델 신경회로망은 사용자의 성향에 더욱 가까워지게 된다.

2. 시스템의 구조

PReA의 전체 구조는 그림 1과 같다. 먼저 미리 입력되어 있는 논문들의 서지 정보 데이터 베이스의 사용 내역을 관찰하여 특성을 추출한다. 이 과정은 그림 1의 부분 중 점선으로 표시된 부분으로써 그림 2에서 더욱 자세하게 나타나 있다. 즉, 사용자가 사용한 질의어, 질의어에 대한 결과, 선택하거나 인쇄 또는 저장한 서지 정보 등이 프로그램 사용 도중에 기록된다. 모든 정보는 단어 수준으로 나뉘어지고 각 단어의 중요도와 관련성을 결정한 뒤에 사용자 모델 신경회로망을 구축하고, 적용시키는데 이용된다.

적용된 사용자 모델 신경회로망에서 중요도가 높은 단어를 추출해서 웹 검색을 위한 질의어를 생성하고 웹을 검색하게 된다. 여기서는 웹을 직접 검색하는 것이 아니라 Altavista의 검색 엔진을 이용하는

메타 검색 방식을 취하였다. 검색된 웹 문서를 단어 별로 나누어서 리스트를 만든다. 이 단어의 리스트가 사용자 모델 신경회로망에 적용되면 단어 리스트에 있는 단어의 부하와 그 단어와 관련도가 높은 단어들의 부하의 합을 구해서 총 에너지를 계산할 수 있게 된다. 에너지가 높을수록 웹 문서가 사용자 모델 신경회로망과 유사하다는 의미이므로 웹문서의 순위를 사용자에게 제시할 수 있게 된다.

순위가 결정된 웹 문서는 사용자가 선택을 하느냐 하지 않느냐에 따라 각각 양과 음의 부하로 사용자 모델 신경회로망을 적용하는데 쓰이고, 선택된 웹 문서의 정보(URL, 제목, 날짜 등)은 서지 정보 데이터 베이스로 저장된다.

이런 방식으로 사용자가 서지정보 데이터베이스를 이용하는 시간이 길고, 웹 문서 검색을 자주 실시하면 할수록 사용자 모델 신경회로망은 사용자의 성향에 더욱 가까워 지게 된다.

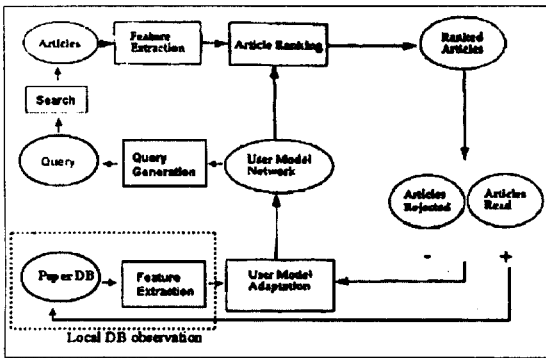


그림 1. PRaE 시스템의 구조

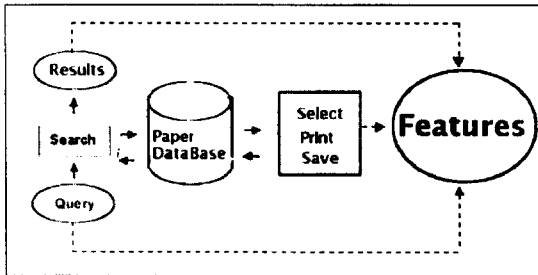


그림 2. Local Database Featuring

3. 사용자 모델 신경회로망

사용자 모델 신경회로망은 Jennings 가 1992년 사용자 중심의 뉴스 필터링 시스템 연구 논문에서 발표한 것으로, 단어들 사이의 관계를 수치적으로 표현해서 사용자에게 중요한 의미가 되는 Concept에 오랜 시간에 걸쳐 적용하게 만들었다.[3]

사용자 모델 신경회로망의 구조는 그림 3과 같이 각 단어들의 부하와 그것들 사이의 부하로 구성되어 있다.

본 연구에서는 기존의 사용자 모델 신경회로망에서 각 단어들의 부하를 조정하는 방법을 리서치 에이전트 시스템에 맞게 향상 시켰다. 즉, 부하의 결정식을 보다 정교하게 만들고, 웹문서의 순위결정 계산

을 보다 단순화 시켰다.

본 연구에서의 회로망은 서지 정보 데이터 베이스와 웹에서 검색된 문서 두 가지를 통해서 구축되고 적용하게 된다.

서지정보 데이터 베이스에서는 사용자가 사용했던 서지 정보를 단어 수준으로 나누어서 단어 리스트를 만들고 다음과 같은 식에 따라서 부하를 계산한다. 단어 i의 부하는 다음과 같다.

$$W_{iNEW} = W_{iOLD} + tf_i \times \alpha^i \times F_k \times A_i$$

W_{iNEW} : 단어 i의 새로운 부하값,

W_{iOLD} : 단어 i의 이전의 부하값,

tf_i : 단어 i가 자주 나타나는 빈도(term frequency) [5]

α^i : 서지 정보 부하 상수 (웹문서 상수 보다는 큰 값)

F_k : 필드 별 상수 (k=저자, 저널명, 제목, 초록 등..)

A_i : 서지 정보 접근 상수 (t=질의어, 검색결과, 인쇄, 저장 등..)

검색된 웹 문서는 단어 수준으로 나누어서 리스트가 만들어지고 다음과 같은 식에 따라서 부하를 계산한다. 단어 i의 부하는 다음과 같다.

$$W_{iNEW} = W_{iOLD} + tf_i \times \alpha^a \times f \times H_k$$

W_{iNEW} : 단어 i의 새로운 부하값,

W_{iOLD} : 단어 i의 이전의 부하값,

tf_i : 단어 i가 나타나는 빈도수(term frequency),

α^a : 웹 문서 부하 상수(서지정보 상수보다는 작은 값),

f : 피드백 상수(선택된 것이면 +1, 아니면 -1),

H_k : Hypertext의 필드별 상수(k=제목, 본문, 강조체 등..)

단어 i와 단어 j 사이의 부하는 다음과 같다.

$$W_{ijNEW} = W_{ijOLD} + C_{ij}$$

W_{ijNEW} : 단어 i와 단어 j 사이의 새로운 부하값,

W_{ijOLD} : 단어 i와 단어 j 사이의 이전의 부하값,

C_{ij} : 단어 i와 단어 j가 같이 나타나는 빈도수,

초기 사용자 모델 신경회로망은 서지정보 데이터 베이스의 사용을 관찰해서 구축을 하고 이후에는 웹 문서와 함께 회로망을 사용자에게 적용시키게 된다.

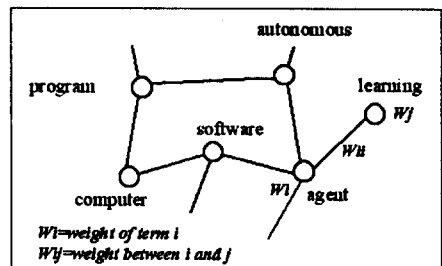


그림 3. 사용자 모델 신경회로망

4. 웹 문서의 순위 결정

사용자 모델 신경회로망에서 추출된 질의어를 통해 검색된 웹문서는 다시 회로망을 통해 중요도 순위를 결정할 수 있다. 먼저 웹 문서에서 단어의 리스트를 뽑아낸다. 그 다음 첫 단어부터 순서대로 각각

의 단어가 회로망에 속해 있는지 본다. 만약 속해 있으면 회로망 속의 단어는 활성화되고 활성화된 단어와 연결되어있는 단어 중에 그 사이의 연결된 부하가 임의의 값보다 크면 연결된 단어도 활성화된다. 이와 같은 과정을 첫 단어부터 끝 단어까지, 그리고 더 이상 활성화되는 단어가 없을 때까지 실행한다. 그리고 활성화된 단어들의 부하 값을 모두 더한 것이 웹 문서와 회로망 사이의 유사도가 된다. 그리고 각 웹 문서마다 유사도를 비교해서 검색된 웹 문서의 순위를 결정할 수 있다. 유사도를 구하는 식은 다음과 같이 표현할 수 있다.

$$e = \sum W_i \quad W_i : \text{활성화된 단어의 부하}$$

이런 방식의 순위 결정은 의미가 유사한 단어의 활성화를 유도한다. 예를들면 "software"와 "computer"만 나타나 있다고 하더라도 회로망의 활성화 과정 속에서 "computer"라는 단어도 활성화되게 된다. 이는 문서 검색 과정에서 정확한 단어 사용에 보다 덜 민감할 수 있는 장점을 가져온다.

5. 구현 및 결론

PREA는 미리 개발된 서지정보 데이터 베이스 검색 프로그램에 사용자 모델 신경회로망 모듈을 접합시켜서 구현하였다. 그림4는 서지정보 데이터 베이스 검색 프로그램의 인터페이스이고, 그림5는 웹 문서를 검색하고 순위를 결정하는 과정을 보여준다. 특히 그림6에서는 기존의 사용자 모델 신경회로망의 부하를 직접 눈으로 확인할 수 없었던 것을 감안해서 사용자가 각 단어의 부하의 크기와 단어들 사이에 연결된 부하의 크기를 상대적으로 눈으로 확인할 수 있도록 하였다.

본 논문에서는 사용자가 논문 서지 데이터 베이스를 사용하는 것을 바탕으로 사용자의 성향을 파악하고 그에 맞는 웹 문서를 검색하고 순위를 정해 사용자에게 제시하는 사용자 중심의 리서치 에이전트 시스템을 구현해 보았다. 이는 자신이 연구하고 있는 논문과 관련되는 웹문서를 찾기 위해 많은 시간을 보내는 사용자들에게 매우 유용할 것으로 보인다. 사용자의 성향을 모델링 하기 위해서 단어들 사이의 관련성으로 이루어진 사용자 모델 신경회로망을 이용하여 정보요구에 대한 의미를 파악할 수 있도록 했다. 앞으로 오랜 기간동안의 실험을 통해 성능을 평가하는 것이 필요하고 단어 부하에 대한 보다 정교하고 동적인 적용 방법의 개발이 요구된다.

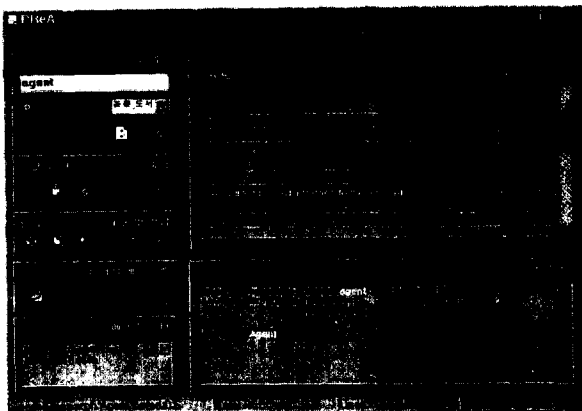


그림 4. PREA의 사용자 인터페이스

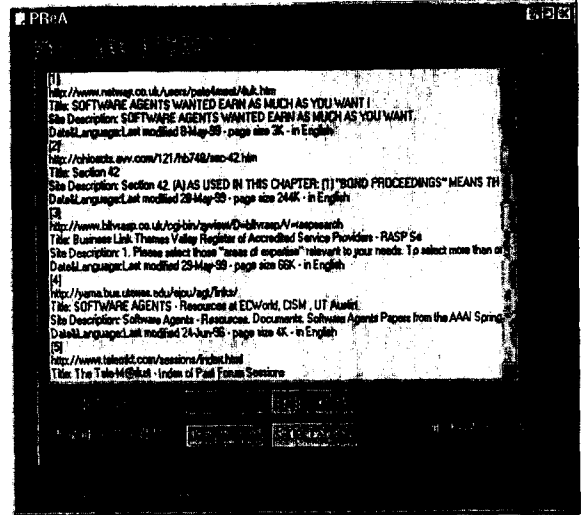


그림 5. 웹 문서 순위 결정

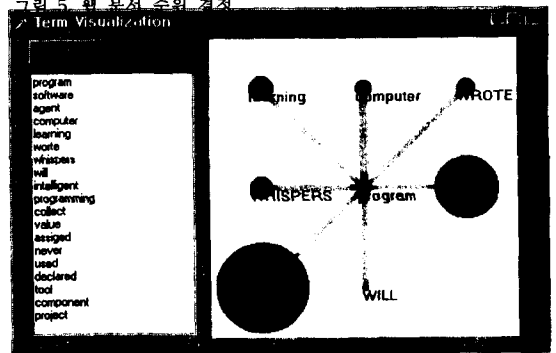


그림 6. 사용자 모델 신경회로망의 Visualization

참고문헌

- [1] Hyacinth S. Nwana, Divine T. Ndumu, "A Perspective on Software Agents Research", *BT Journal*, London, BT Journal, 1998.
- [2] Maes, P., "Agents that Reduce Work and Information Overload", *Communications of the ACM*, vol.37, No.7 pp.40, 1994.
- [3] S J Soltysiak, I B Crabtree, "Automatic learning of user profiles - towards the personalisation of agent service", *BT Techno Journal*, Vol.16, No. 3, 1998.
- [4] 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
- [5] G. Salton, "Automatic Text Processing", Addison-Wesley,