

웹 에이전트에서 행위 기반 학습에 근거한 사용자 관심도 예측에 관한 연구

신 민 수°, 박 재 균 , 이 창 훈
건국 대학교 컴퓨터 공학과

User Concern-degree Predict System Using the User Action Learning of Web Agent

Min-su Shin°, Jae-kyun Park , Chang-hun Lee
Dept. of Computer Engineering, KonKuk University

요 약

최근 인터넷의 사용이 폭증하며 기존 검색 엔진의 단점을 극복하고자 웹 에이전트 연구가 활성화되고 있다. 이에 웹 에이전트의 구조에서 사용자의 관심, 비관심 문서의 분류 수집에 있어 직접적인 피드백에 의한 학습은 사용자에 부담과 불편을 제거하고자 간접적인 사용자 행위 분석에 의한 성향 학습 시스템을 제안한다.

1. 서 론

최근 몇 년 간의 인터넷 사용자 수를 보면 거의 기하급수적으로 늘어나고 있음을 알 수 있다. 그럼 과연 무엇 때문에 이렇게 사용자의 수가 늘어나고 있는 것일까?

간단히 말하자면 정보 때문이다. 인터넷은 정보의 보고라 불리울 만큼, 수 많은 정보 자원이 세계 곳곳에 산재 되어 있고 이를 인터넷을 통해 하나로 묶어 줄 수 있기 때문이다. 이런 인터넷에는 하루에도 수 많은 정보들이 새로이 인터넷에 부가 되어진다. 여기서 주목할 사실은 날마다 방대해지는 인터넷 자원 속에서 사용자 자신이 원하는 정보를 찾아 내기란 그 만큼 더욱 어려워지고 있다는 것이다. 이렇게 자신이 원하는 정보를 찾아내기 위해 우리는 야후(<http://www.yahoo.com>), 알타비스타(<http://www.altavista.com>), 심마니(<http://www.simmani.com>)와 같은 검색 엔진(Search Engine)을 사용한다.

검색 엔진은 사용자가 원하는 정보를 찾기 위해 원하는 검색어(Key Word)를 검색 엔진의 질어어 창에 입력하면, 그 검색어에 의한 검색 결과의 링크들을 다시 보여주게 된다. 이런 과정에서 사용자가 원하는 정

보를 위해 매년 검색어를 입력하는 일은 여간 번거로운 일이 아닐 수 없다. 또한 검색되어진 결과가 모두 사용자의 관심 대상 일 수도 없는 일이다.

이런 검색 엔진의 단점을 보완 하고 사용자의 편의성을 증대 시키기 위해 검색 엔진의 개량화는 계속 시도되어졌다. 하지만 근본적으로 사용자의 수동적인 검색 위주로 그 한계가 있다.

이에 능동적인 사용자 성향에 따른 정보 검색을 위한 연구로 에이전트(Agent)를 이용한 연구가 활성화 되었다. 이와 같이 웹 검색에 에이전트를 활용하여 웹 에이전트(Web Agent)란 분야가 개척되었는데, 웹 에이전트는 지능을 가지고 스스로 행동할 수 있는 시스템으로 인터넷 상의 링크들을 검색하는데 있어 사용자의 취향에 비추어 관심 있는 문서들을 검색하고 이 검색 결과를 다시 선별하여 사용자에게 제시해 줄 수 있는 시스템을 말한다.

본 연구에서는 이런 전체적인 웹 에이전트의 동작에서 사용자의 취향(Profile)- 사용자에게 관심 또는 비관심 문서인지를 판단하기 위해 사용자의 성향을 학습하여 구성한 자료 -을 위해 웹을 검색하는 사용자의 부가적인 행동을 관찰하고 관찰된 결과를 분석

하여 그 문서가 사용자에게 관심 또는 비관심 문서인지를 판단할 수 있는 사용자의 성향을 예측할 수 있음을 보이고자 한다. 물론 이후의 단계인 사용자 관심에 의해 선택 되어진 문서에 대해서 키워드를 추출하여 사용자의 학습 프로파일을 구축하는 문제도 중요한 부분이지만 본 논문에서는 제외 시키기로 하겠다.

이런 사용자의 관심도 성향을 학습하는 방법에 있어 직접적인 피드백(Feedback)을 통해서 사용자가 검색한 문서에 대한 선호도를 판별할 수 있고, 또 간접적으로 사용자의 행위를 모니터링 하여 판별할 수도 있다. 하지만 간접적인 방법에 의한 학습에 대해서는 아직까지 연구가 많이 부족한 상태이다.

이런 기존의 웹 에이전트 시스템에서 사용자의 관심, 비관심 문서를 판별하는 부분에 대한 가장 큰 단점으로 생각할 수 있는 것은 사용자에 대한 직접적인 피드백(Feedback)을 요구한다는 것이다. 이런 단점을 극복하고자 본 연구에서는 간접적인 학습 방법에 의한 사용자의 행위를 관찰하고 이를 분석함으로써 직접적인 피드백에 의한 사용자의 불편과 부담을 덜어 보고자 행위 분석 학습에 의한 에이전트 시스템을 제기한다. 이 시스템에는 크게 두 부분으로 구성되는데 사용자의 행위를 관찰하고 이를 분석할 수 있는 에이전트 시스템으로 사용자의 웹 상에서 검색 행위나 기타적인 입력 사항을 분석하여 사용자의 관심을 간접적으로 추론하는 부분과 이렇게 분석된 결과를 이용하여 사용자의 성향을 학습하여 추후로 사용자의 행위를 바탕으로 추가적인 갱신이 가능하도록 하는 시스템이다.

다음으로 본 논문에서 각 장에서 언급하고자 하는 내용을 살펴보면, 1장은 위의 서론 부분 이고, 2장은 관련 연구로 에이전트의 전체적인 설명에서부터 웹 에이전트까지 그리고 이런 웹 에이전트의 사례 분석과 학습 방법의 하나로 선정한 신경망의 역전파 알고리즘에 대해 설명한다. 3장에서는 시스템의 설계와 구현에 대해서, 4장에서는 구현된 시스템에 의한 실험과 그 결과를 제시하고 5장에서는 실험을 통한 결론과 더불어 향후 과제에 대해서 언급한다.

2. 관련 연구

2.1 에이전트

1980년대 초반부터 언급되기 시작한 에이전트(Agent)의 개념은 인공지능(Artificial Intelligent)분야의 하나로, Object that thinking, 즉 자신 스스로의 지능을 가지고 행동할 수 있는 소프트웨어의 단위를 뜻하며 기존의 기능 위주의 프로그램과는 달리 에이전트는

목적 지향적인 프로그램이다.

에이전트의 분류는 기준을 어느 특성에 두느냐에 따라 다중 에이전트(Multi-agent), 지능 에이전트(Intelligent agent), 이동 에이전트(Mobile agent), 보조 에이전트(Assistant agent)등 그 분류는 매우 다양하다. 본 논문에서는 웹 검색 엔진에 관점을 두었기에 검색 에이전트(Information/Internet Agent)에 대해 언급하고자 한다.

인터넷 에이전트는 흔히 정보 검색 에이전트로 언급하는데, 우리가 관심을 두고 있는 웹 에이전트와 초기의 WebRot등이 여기에 해당된다. WebRot은 각 웹 페이지에서 링크를 따라서 세계 곳곳에 분산된 네트워크를 향해 하며 문서를 검색하여 정보를 모아오는 기능을 한다. 이러한 정보 검색을 통해 방문한 사이트나 새로운 링크 등을 사용자에게 알려준다.

에이전트는 사용자의 행위를 대신 수행하는 프로그램이다. 하지만 스스로의 지능에 의해 주어진 상황에 능동적으로 행동 할 수 있어야 한다. 즉 주어진 지식을 스스로 학습하여 발전 시켜 나가는 목적 지향적인 행위와 환경의 변화에 적절한 행동을 할 수 있다는 것이다.

본 연구에서 에이전트를 이용하는 것은 사용자의 행위를 기반으로 관심도를 판단하고 나아가서는 이 관심도를 이용해 검색을 수행하기 위하여 사용자 관심의 변화를 파악 할 수 있어야 한다.

2.2 신경망

분산처리 시스템(Parallel Distributed Processing System)으로서 정보의 저장, 처리 및 전달을 Neural Network내의 특정 부위들이 분담하는 것이 아니라 항상 Neural Network 전체로 그 기능을 수행한다. 그 특징은 다음과 같다.

첫째, Neural Network은 자기 조직화 시스템(Self Organizing System)이다. 즉 Neural Network은 새로운 이미지나 패턴 또는 사례가 주어졌을 경우 이를 기억하기 위해 자동적으로 자신의 내부 상태를 바꾼다. 학습된 이미지나 패턴의 일부 또는 손상된 정보로부터 완전하고도 정확한 정보를 재생해 낼 수 있는 능력이 있다. 이와 같은 기능을 연상기억(Associative)이라고 한다. 둘째, Neural Network에서 학습된 지식은 뉴런들간의 연결 Weight에 의해 분산 저장된다. 학습은 가장 중요하면서도 가장 기본적인 요소로서 Neural Network를 사용자가 원하는 상황에 적응하도록 훈련시킬 수 있음을 의미한다. 이와 같이 분산 저장된 학습은 Neural Network의 일부가 손상되어도 본래의 기능을 수행 하는 데에는 큰 지장을 받지 않

게 되는 Fault Tolerance기능을 갖게 한다. 셋째, Neural Network는 사용자의 측면에서 볼 때 기존의 컴퓨터처럼 복잡한 프로그래밍 과정을 요구하지 않는다. 이용자는 신경회로망의 구조 및 학습 방법을 지정함으로써 간단히 원하는 결과를 얻을 수 있으며, 내부 처리 절차 또한 매우 간단히 이루어진다. 넷째, Neural Network는 패턴의 분류나 인식에 있어 강력한 힘을 발휘하며, 우리가 부딪히는 대다수의 문제들이 패턴 분류의 문제임을 볼 때 폭 넓은 적용 범위를 지닌다.

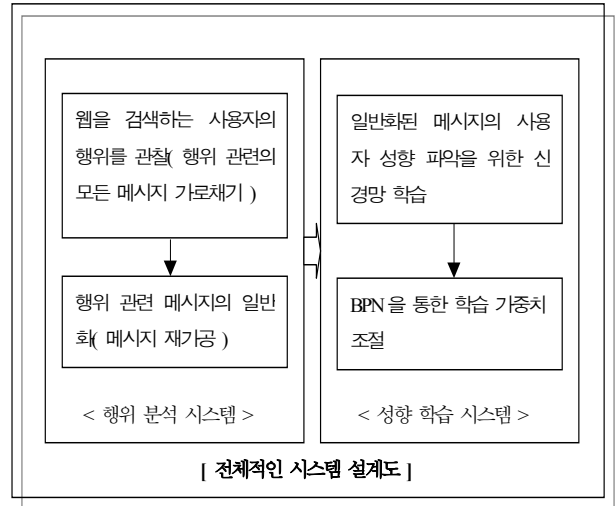
출력값을 결정하는 가장 중요한 요소는 연결 가중치이다. 각 처리 요소들의 출력값들은 결국 신경회로망의 출력 결과를 결정하게 되므로 신경 회로망을 이용하여 원하는 출력값을 얻기 위해서는 연결 가중치를 조절하여야 한다. 모든 신경회로망은 주어진 사례를 이용하여 처리 요소들 사이의 연결 가중치를 조정하게 된다. 본 논문에서는 역전파(Back Propagation Network)을 이 가중치의 조절을 위해 사용한다. 학습 방법은 다음과 같다.

- 1) 새로운 학습 패턴에서 입력 패턴을 Back Propagation Network의 입력층에 입력시킨다.
- 2) 입력된 입력값들을 이용하여 입력층에서 은닉층(Hidden Layer)으로, 또 은닉층에서 다시 출력층, 또는 다음 은닉층으로 신호를 전파한다. 이러한 신호 전달 과정은 출력층에서 최종 출력을 내보낼 때까지 진행된다.
- 3) 출력층에서 출력된 출력치들과 학습패턴에서의 실제 또는 원하는 값과의 오차를 계산한다. 이 때 오차가 정의된 임계치 이하로 떨어지면 학습 과정을 종료한다.
- 4) 계산된 오차를 이용하여 2)항에서 역순으로 즉, 출력층에서 은닉층으로, 은닉층에서 입력층으로의 연결 강도와 고정 입력치들을 수정한다.
- 5) 현재의 학습 패턴 대신 또 다른 학습 패턴을 선정하여 다시 1)항부터 절차를 진행한다

3. 행위 기반 예측 시스템 설계 및 구현

전체적인 시스템의 설계는 크게 두 개의 시스템으로 다시 분류 될 수 있다. 먼저 사용자 행위의 일반화를 통한 사용자 행위 분석 시스템과 이로 인해 분석된 정보에 의해 사용자 성향을 학습하는 성향 학습 시스템이다.

위 그림에서 전체적인 시스템 설계도를 보여준다. 행위 분석 시스템에는 크게 세 부분으로 구성되는데 첫째는 웹을 검색할 때 나타날 수 있는 사용자의 모든 행위의 모니터링이다. 이 행위들은 두 번째 단계에

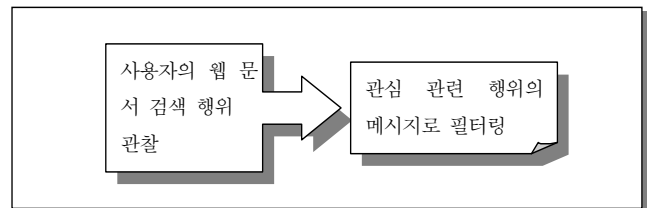


서 윈도우에서 발생할 수 있는 모든 메시지들을 가로채어 관찰 할 수가 있다. 하지만 모든 메시지가 관심 있는 문서를 검색하는 사용자의 행위는 아니므로 이런 행위 관련 메시지의 노이즈(Noise)를 필터링 하여야 한다. 이런 필터링을 통해서 사용자 행위의 일반화를 즉, 메시지들의 의미를 재가공 할 수 있는 것이다. 이런 메시지들의 재가공을 통해서 학습에 필요한 정보들을 추출하는 단계가 행위 분석 시스템의 궁극적인 목표이다.

성향 학습 시스템에선 이런 행위 정보를 입력 노드로 신경망 학습을 하게 된다. 특히 적절한 가중치의 적용을 위해 역전파 알고리즘을 사용하여 중요도가 높은 행위를 더욱 부각시키고 반대의 행위는 낮춰지게 된다. 이렇게 함으로써 사용자의 성향이 올바르게 파악되어 실제적 문서의 관심, 비관심을 판단하는데 정확성을 높일 수 있다.

3.1 행위 분석 시스템

행위 분석 시스템을 간단히 설명하면, 웹을 검색하는 사용자의 행위를 모니터링 하고 이와 관련된 행위 정보를 수집하고, 이 정보를 분석하는 시스템이다.



위의 그림처럼 두 단계로 설계되며 각각의 단계에선 다음과 같은 작업을 수행하게 된다.

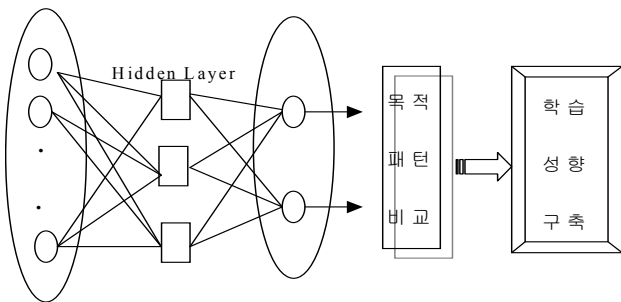
첫째, 사용자의 문서 검색 행위를 모니터링 하는 것인데 이전에 연구에서는 대개 간접적인 모니터링이라 하여 사용자의 특정 행위의 관찰을 위해 간이 검색 브라우저를 제작하여 수행하였는데 본 논문에서는 시스템 차원에서 즉, WIN32기반의 윈도우 환경에서 오고가는 메시지를 가로채어 확인한다. 그리고 이렇게 가로채어 온 모든 메시지가 사용자의 중요 행위인 것은 아니다. 다시 말하면 의미 없는 메시지들에 대한 필터링(filtering)이 필요하다. 이 과정은 두 번째에서 이루어지는데 웹을 검색하는 사용자의 행위를 일반화시키기 위함이다. 이를 행위 분석 구성 요소라 정의하고 본 논문의 실험을 위해 다음처럼 나타낸다.

- PB print button, 문서의 인쇄
- BB-bookmark button, 즐겨 찾기에 추가
- SB-save button, 문서의 저장
- LBD-left button down, 왼쪽 마우스 버튼 클릭
- RBD-right button down, 오른쪽 마우스 버튼 클릭
- MM-mouse move, 마우스 이동
- SCB-scroll bar, 스크롤
- T-time, 문서 검색 시간
- KD-key down, 키 입력
- KC-key char down, 키 위치 이동

결국, 이 과정을 통해서 사용자 행위 구성 요소에 따른 분석을 수행하여 이와 관련된 메시지만을 추출하게 되는 것이고 이를 통해 사용자가 보는 문서의 관심도를 예측하게 되는 것이다.

3.2 성향 학습 시스템

성향 학습 시스템에선 행위 분석 시스템을 통해 얻어진 사용자 행위 분석의 결과 즉, 사용자가 관심 있는 문서를 검색할 때 취하는 행위 성향을 학습하게 된다.



이런 학습을 위해 위의 그림처럼 본 연구에서는 신경망을 이용했고 중요 행위의 부각을 쉽게 하기 위해 학습 가중치(weight)조절을 역전파 알고리즘을 사용해 조절한다.

행위 분석 구성요소들을 입력 노드가 되어 임의의 가중치에 의해 출력 노드를 산출한다. 우리가 원하는

목적 패턴은 관심=1, 비관심=0이 되고 만족이면 (1,0), 불만족이면(0,1) 이를 만족시키기 위한 근사 수치를 찾기 위해 Back Propagation을 이용 계속해서 가중치를 조절한다. 이런 과정에서 목적 패턴에 영향을 많이 미치는 입력 노드는 그 만큼 더 중요한 행위로 부각이 되는 것이다. 한 문서 내지는 여러 다른 문서를 검색하는 사용자의 행위를 계속 학습 시켜서 개인 사용자에 대한 성향을 학습시킬 수 가 있게 된다.

4. 실험 및 결과분석

	학습문서	실험문서	관심문서
경우 1	20	20	3
경우 2	50		8
경우 3	70		11
경우 4	90		15
경우 5	100		11

첫 번째 실험은 학습 대상의 문서 수를 실험 인자로 하여 사용자의 만족도 측정이었다. 실험에 의해 학습 문서의 수가 증가 할수록 사용자의 만족도를 높일 수 있다. 하지만 한계 수치를 넘으면 오히려 줄어든다는 것을 알 수 있다.

	관심문서 (실제관심문서)	관심문서 (실제비관심문서)	비관심문서 (실제관심문서)
사용자(%)	69	17	14

두 번째 실험은 문서 100개를 가지고 전체 문서 수에 대한 각 항목에 대한 백분율을 구해 보았다. 학습율은 0.7, 학습 반복 회수는 20회이다.

두 실험에서 사용자 관심도 예측에 있어 생각보다 어느 정도의 정확도를 적중했는데 오차는 가중치 조절에 원인이 있는 듯 하다.

5. 결론 및 향후과제

웹 에이전트의 연구는 이미 많이 진행되고 있으나 아직은 직접적인 피드백에 의한 학습으로 비교적 정확한 예측은 될 수 있으나 사용자의 불편이나 부담 또한 크다. 이에 본 연구에서는 시스템 차원에서의 사용자 행위 분석을 통한 간접적인 학습 방법에 의한 사용자 관

심도의 예측을 제안했으나 시스템 차원에서 바라본 사용자 행위에 대한 일반화가 명쾌하지 못한 듯 하다. 각 개인차에 의한 행위의 일반화가 다소 어렵기 때문이다.

이런 행위의 일반화를 구조적으로 잘 정의하여 최적화 할 수 있다면 더욱 뚜렷한 결과를 볼 수 있으리라 생각하고 이런 이유로 직접적인 피드백에 의한 방식에 비해 다소 떨어지는 성능을 나타낸다.

하지만 사용자는 웹 에이전트로부터 독립적이기에 별 다른 신경을 쓰지 않아도 된다. 현재 많은 어플리케이션 프로그램의 개발 추세가 사용자의 편의성을 극대화시키는 방향으로 진행되기에, 눈에 나타나는 단점을 보완한다면 상업적인 효용성도 클 것 같다.

참고 문헌

- [1] William B. Frakes, Richard Baeza-Yates, "Information Retrieval Data structures & algorithms"
- [2] B. Markom and S. Yoav, "Learning Information Retrieval Agents: Experiments and Hypertext", Carnegie Mellon University 1995
- [3] J. Rocchio, "Relevance Feedback in Information Retrieval", The SMART Retrieval System: Experiments in Automatic Document Processing
- [4] C. Alper and H. Colin, "Agent Source Book". Wiley Computer pub, 1997
- [5] Mladenic, D. (1998) Turning Yahoo into an Automatic Web-Page Classifier Proceedings of the 13th European Conference on Artificial Intelligence ECAI'98