

웹 상에서 사용자 브라우징 행위를 이용한 하이퍼텍스트 네트워크 재구성 (Hypertext Networks Restructure using User Browsing Behaviors on WWW)

백 영 태* 이 세 훈**
(Yeong-Tae Baek) (Se-Hoon Lee)

요 약

이 연구는 사용자의 브라우징 행위에 따라 자동적으로 하이퍼텍스트 네트워크를 재구성하는 세 가지 학습 규칙을 제안하고, 실험적 시스템을 구현하였다. 시스템은 링크 가중치를 하이퍼링크 네트워크에 부여하고 학습 규칙에 따라 가중치를 변경한다. 학습 규칙은 하이퍼링크가 얼마나 자주 이용되고 있는지에 따라 해당 하이퍼링크의 가중치만 변경되며, 다른 하이퍼링크에는 영향을 미치지 않는다. 네트워크 구조의 변경은 링크 가중치의 내림차순에 따라 동적으로 링크가 배열되어 사용자에게 제시된다. 이것은 협력 필터링 기술의 장점과 탐색 지원 접근 방식을 혼합한 것이다. 실험을 위해 임의적인 하이퍼텍스트 네트워크를 만들고 사용자의 브라우징 선호에 따라 네트워크 구조가 변화되는 과정을 관찰한다.

ABSTRACT

We have suggested three learning rules and implemented an experimental system that automatically restructures hypertext networks according to their user browsing behavior. The system applies link weights to the hypertext networks and updates these weights according to the learning rules. The learning rules are based on how often a particular hyperlink is being traversed and operate on strictly local information of link traversals. Changes in network structure are fed back to users by dynamic link ordering according to descending link weight. This approach is used collaborative filtering and navigation support techniques. The system has been investigation to be able to structure random hypertext networks into valid representations of their user browsing preferences in experiments.

1. 서론

인터넷과 월드와이드웹의 정보는 다른 매체가 수십 년, 수 백년동안 축적한 것들을 거의 모두 담을 정도로 증가되어 왔으며, 이에 따라 많은 사람들이 인터넷에서 정보를 제공하거나 찾고 있다[2, 13, 14]. 그러나 웹 상의 정보는 페이지의 수와 링크의 수가

증가됨에 따라 사용자는 원하는 정보를 찾는데 더욱 어려움을 겪고 있다[12, 13, 14]. 데이터베이스나 다른 정보 저장소로부터 정보를 검색하는 것은 정보 검색 분야에서 연구가 있어왔으나 데이터베이스 중심의 질 의기반 검색에 대한 연구로서, 웹이 갖고 있는 사실상의 하이퍼텍스트 구조가 무시되었고 단지 연결되지 않은 전자 문서의 저장소로만 다루어진 것이 많다.

* 정회원 : 김포대학 컴퓨터계열 조교수

** 정회원 : 인하공업전문대학 전자계산기과 부교수

논문접수 : 2001. 10. 21.

심사완료 : 2001. 11. 4.

웹은 정보 저장 수단으로서 뿐만 아니라 정보의 가치를 높을 수 있는 정보의 네트워크 구조가 가능하다. 이러한 특성은 사용자들이 네트워크 구조를 쉽게 브라우징하고 문서 사이에 연결을 따라 가기만 하면, 원하는 정보들을 쉽게 얻을 수 있다. 웹의 정보 검색에서 발생하는 상호작용 방식을 전통적인 정보검색 기술로는 부족한 점이 매우 많다[4]. 웹이 점점 더 중요한 정보 매체가 될수록 하이퍼텍스트 네트워크에서 사람들의 탐색과 검색의 효율성을 증진하기 위한 연구는 중요한 의미를 갖는다[2, 11, 14].

브라우징하는데 사용자의 어려움을 해결하기 위한 방법들을 제안하고 있다[3, 7]. 즉 기본적인 네트워크 구조에 관해 사용자에게 정보를 제공하는데 시각화, 그래픽맵, 안내 등의 방법을 이용한다. 또한 여러 가지 학습 결과가 특정 하이퍼시스템에 맞서 사용자를 도울 수 있다고 소개되었다[8]. 그러나 웹에서 검색 중심의 문제들이 불완전한 사용자 지식 때문이라고만 할 수는 없으며, 여러 요소들 사이에 매우 복잡한 상호 작용 즉, 설계자의 선호, 사용자의 목적 및 관심, 특정한 인터페이스 등이 포함되어 있다[10].

가장 중요한 요소 중에 하나는 웹의 정적인 면과 부적절한 하이퍼텍스트 구조이다. 이러한 문제를 해결하기 위한 연구로서, 협력 필터링(collaborative filtering) 기술을 이용한 시스템 등이 있다. 그러나 이러한 시스템들도 웹에서 정보를 찾는데 공통적인 브라우징 같은 상호작용 정보 전략을 지원하는데 사용되기는 어렵다. 다른 접근 방식으로 [7]에서는 사용자가 브라우징 하는 과정을 추적하고, 추적할 의미가 있는 링크를 기록한다. 이러한 기록을 이용해 다른 사용자에게 추천한 결과 좋은 결과를 얻었다. 다른 예로, [1]에서는 협력 필터링과 브라우징 지원의 장점을 혼합하였다. [9]에서는 사용자 기호에 관한 정보를 저장하는데 연상 매트릭스(associative matrices)를 사용하는 것이 상당한 효과가 있음을 보여주고 있다. 그러나 이러한 시스템들은 웹을 위해서 설계되지 않았기 때문에 웹의 특징적인 부분을 충분히 적용하고 있지 못하다.

따라서 이 연구에서는 기존 하이퍼텍스트 네트워크의 단점을 줄이기 위한 기능들을 구현하기보다는, 하이퍼텍스트 네트워크 구조 자체를 변경하는 규칙 모델의 제안과 이러한 모델을 반영한 시스템을 개발하는 접근 방식으로 진행한다. 시스템은 하이퍼텍스

트 네트워크를 링크에 대한 사용자 선호 패턴 그룹을 사용자 모델로 구조화한다. 이것은 협력 필터링 기술의 장점과 탐색 지원 접근 방식을 혼합한 것이다.

2. 동적 하이퍼텍스트와 협력 필터링 기술

이 장에서는 동적 하이퍼텍스트 구축 방법과 기술을 고찰하고 협력 필터링 기술의 관련 기술들을 분석한다.

2.1 동적 하이퍼텍스트

일반화의 어떤 단계에서 하이퍼텍스트는 링크에 의해 연결된 노드나 하이퍼문서인 페이지의 집합으로 구성된다. 각 페이지는 어떤 지역적 정보와 관련 있는 페이지에 대한 몇 개의 링크를 가진다. 하이퍼텍스트 시스템은 모든 접근 가능한 페이지에 대한 링크를 제공하는 전체 사이트 맵과 색인을 포함할 수 있다[3, 5].

동적 하이퍼텍스트에서 동적으로 될 수 있는 것은 내용 단계로 일반 페이지의 내용과 링크 단계로 일반 페이지로부터의 링크, 색인 페이지, 그리고 사이트 맵 등이 있다. 하이퍼텍스트의 동적인 부분은 내용 단계와 링크 단계의 서로 다른 두 개의 부류로 구분하고 이를 각각 동적 프레젠테이션과 동적 탐색 지원이라고 부른다.

동적 프레젠테이션 기술의 아이디어는 현재 사용자의 지식, 목적, 사용자의 다른 특징들을 가지고 사용자가 접근하려는 페이지의 내용을 동적으로 변화시키는 것이다. 예를 들어, 초보자에게는 추가적인 설명을 제공하는 반면에 허가받은 사용자에게는 보다 자세하고 충분한 정보를 제공한다. 하이퍼텍스트 시스템에서 일반 페이지의 내용은 고전적인 하이퍼텍스트 시스템과 같이 텍스트로만 구성되는 것이 아니라 다양한 멀티미디어 항목도 포함된다. 이러한 관점에서 하이퍼텍스트 시스템은 동적 텍스트 프레젠테이션, 동적 멀티미디어 프레젠테이션으로 구분한다. 그러나 하이퍼텍스트에서 동적 프레젠테이션의 모든 작업은 실제로 동적 텍스트 프레젠테이션으로 이루어지고 있다. 동적 프레젠테이션 방법으로는 추

가설명, 선행설명, 비교설명, 설명 방법의 변화, 재분류 등이 있다. 이 중에서 가장 호응이 좋은 방법은 추가적인 설명이 나오는 방법이다. 이 방법의 목적은 이러한 것을 원하거나 필요로 하는 사용자들에게 추가적인 정보, 설명, 실례 등을 제공하는 것이다. 사용자들은 간결하고 기본적인 정보만을 보다 선호하기 때문에 시스템은 어려운 수준의 설명이 있거나 그런 것을 원하지 않는 사용자에게는 설명을 숨긴다. 동적 프레젠테이션에 사용되는 기술로는 조건텍스트, 텍스트펼치기, 프로그래먼트 다양화, 페이지 다양화, 프레임기반 기술 등이 있다.

동적 탐색 지원 기법의 아이디어는 사용자의 목적과 지식, 그리고 개별적인 사용자의 다른 특징들을 가지고 프레젠테이션 되는 링크의 방법을 동적 적용함으로써 하이퍼공간 내의 경로를 사용자가 찾을 수 있도록 도와주는 것이다. 사용자들은 그들이 원하는 정보에 관한 목적을 갖고 있으며, 검색은 더 좋아하게 되거나 단지 필요한 정보를 찾는 하이퍼텍스트 응용에서 안내가 지원될 수 있다. 동적 탐색 지원 기법의 기술로는 방향안내, 링크분류, 링크숨기기, 링크주석, 사이트맵 등이 있으며, 방법으로는 전역적으로 탐색을 지원하는 것과 지역적으로 탐색을 지원하는 방법 등이 있다.

2.2 협력 필터링 기술

협력 필터링은 다른 사용자들의 평가를 기반으로 하여 사용자에게 추천을 생성해 내는 기술이다[6]. 협력 필터링 시스템은 어떤 정보를 이미 보거나 경험한 사람들의 행동과 의견을 가지고 그 정보를 아직 보지 못한 사람들에게 그 정보의 가치를 예측하여 주는 시스템이다. 협력 필터링은 다른 사람들의 평가를 의미적으로 수집하고 분석하여 정보를 찾는 시간을 줄일 수 있다. 협력 필터링 시스템이 없다면, 정보에 대한 평가는 무계획적이고 부정확한 구전에 의해 전달되어 질 수밖에 없을 것이다. 보다 정형화된 평가에 의해, 협력 필터링은 많은 사람들 사이에 일관되고 정확한 평가를 전달하는 것이 가능하게 한다.

협력 필터링 시스템은 사용자에게 추천을 제공하기 위하여 일련의 초기 평가 항목을 필요로 한다. 각 사용자는 추천이 시작되기 전에 평가를 위해 10-15 개의 항목에 응답하게 된다. 시스템은 응답을

저장하고 유사한 선호를 갖는 사람들을 찾아내어서 그들의 평가와 비교한다. 유사한 선호를 갖는 사람들의 집합인 참조 그룹(reference group)을 찾아낸 후, 시스템은 추천 받을 사용자가 아직 보지 않은 항목들을 찾는다. 다음으로 참조 그룹에 있는 사람들의 평가에 기반하여 각 항목에 대한 사용자가 기대하는 평가 점수를 계산한다. 이러한 전체적인 추천 과정은 대부분의 협력 필터링 시스템에서 참조 그룹을 찾는 구체적인 알고리즘과 평가 점수를 계산하는 알고리즘 등이 다를 뿐 동일하다.

협력 필터링 시스템은 데이터베이스에서 사용자들의 평가로부터 특정 사용자의 평가를 예측하는데 사용된다[6]. 많은 알고리즘이 발표되었으며, 이러한 협력 필터링 알고리즘을 메모리 기반 알고리즘과 모델 기반 알고리즘을 구별할 수 있다.

메모리 기반 알고리즘은 추천 받을 사용자에 대한 부분적 정보와 기존 사용자 데이터베이스로부터 계산된 가중치 집합에 기반하여 특정 사용자의 평가를 예측한다. 사용자 i 의 항목 j 에 대한 예측된 평가 \hat{e}_{ij} 는 다른 사용자들의 평가에 대한 가중치 합이다.

$$\text{Estimated evaluation : } \hat{e}_{ij} = ef(s_{i, \cdot}, e_{\cdot, j})$$

$$\text{Estimated function : } ef_{ij}(s_{i, \cdot}, e_{\cdot, j}) = \sum_{t \neq i} e_{jt} w_f(s_{i, t})$$

where, $w_f(s_{i, t})$: weight function

e_{ij} : Evaluation by i^{th} user on j^{th} item

$e_{i, \cdot}$: All evaluations by i^{th} user ($= e_{i1}, e_{i2}, \dots, e_{im}$)

$e_{\cdot, j}$: All users' evaluations on j^{th} item
($= e_{1j}, e_{2j}, \dots, e_{2n}$)

s_{ab} : Degree of similarity of user a and user b

$s_{ab} = sf(e_{a, \cdot}, e_{b, \cdot})$

다음 두 가지 방법 즉 상호관계성식과 벡터유사성식이 두 사용자 사이에 유사성을 계산하는데 가장 널리 사용된다.

Correlation :

$$sf(e_{a, \cdot}, e_{b, \cdot}) = \frac{\sum_{j=1}^m (e_{aj} - \bar{e}_{a, \cdot})(e_{bj} - \bar{e}_{b, \cdot})}{\sqrt{\sum_{j=1}^m (e_{aj} - \bar{e}_{a, \cdot})^2} \sqrt{\sum_{j=1}^m (e_{bj} - \bar{e}_{b, \cdot})^2}}$$

Vector similarity :

$$sf(e_{a, \cdot}, e_{b, \cdot}) = |E_b - E_a|$$

where $E_a = (e_{a1}, e_{a2}, \dots, e_{am})$ and $E_b = (e_{b1}, e_{b2}, \dots, e_{bm})$

모델 기반 기법은 수학적 모델에 기반을 두고 각 항목에 관한 사용자의 평가를 예측한다. 비록 모델 기반 알고리즘에 많은 변수가 있어도 두 가지 그룹 즉 클러스터 모델과 베이시안 네트워크 모델로 분류할 수 있다.

클러스터 모델은 사용자들을 어떤 수의 클래스들로 분류하고 클래스 멤버가 항목 j 에 평가 선호 점수 k 를 줄 확률을 계산한다. 여기서 확률은 표준적인 베이시안 확률 공식을 이용하여 계산할 수 있다.

Bayesian probability:

$$\Pr(C = c, e_1, \dots, e_n) = \Pr(C = c) \prod_{i=1}^n \Pr(e_i | C = c)$$

where, e_i : evaluation for item i

C : class membership

클래스 멤버십 확률 ($\Pr(C = c)$)와 주어진 클래스 ($\Pr(e_i | C = c)$) 평가의 조건적 확률은 사용자 평가의 집합으로부터 계산된다.

베이시안 네트워크 모델은 노드로서 항목을 나타내고 노드의 상태로서 평가를 나타낸다. 이 학습 알고리즘은 각 항목의 의존성에 대한 여러 가지 다양한 모델 구조를 찾을 수 있다. 결과 네트워크에서 각 항목은 평가의 최고 예측 값인 상위 항목의 집합을 가질 것이다.

3. 하이퍼텍스트 재구성을 위한 학습 규칙 모델

이 장에서는 재구성을 위한 세 가지의 학습 규칙 모델의 이론적인 배경과 구체적인 원칙을 설계한다.

3.1 모델의 개요

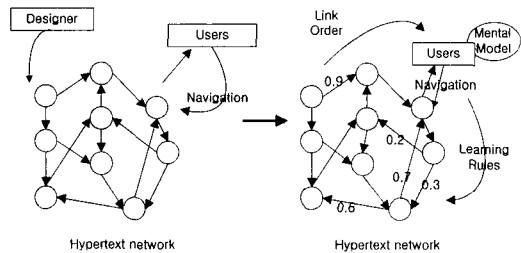
이 연구에서는 하이퍼텍스트의 동적 재구성을 위한 모델을 소개하는데 이는 설계자로부터의 간섭을 거의 받지 않고 탐색되는 동안 자신의 링크 구조를 재구성하는 하이퍼텍스트 모델이며, Bollen의 모델과 많은 아이디어를 공유하고 있다. 모델은 [그림 1]과 같이 연결 강도를 구현함으로써 사용자와 하이퍼텍스트 네트워크간의 피드백 루프를 구성한다.

a) 하이퍼링크에 연결 강도를 할당한다.

b) 일련의 학습 규칙이 사용자의 탐색 결정에 따라 연결 가중치를 변경한다.

c) 네트워크 구조 변경은 연결 강도에 따른 링크의 순서대로 사용자에게 피드백 된다.

[그림 1]은 제안된 모델의 주요 기능의 간단한 윤곽으로 상세한 설명은 다음절에서 한다.



[그림 1] 설계자와 사용자 사이의 상호 작용

[Fig. 1] The adaptive hypertext system closes the feedback loop between designer and users

하이퍼링크는 방향성을 가지며 부울린 형이다. 즉, 하이퍼링크는 하나의 페이지로부터 다른 페이지를 가리키며 하이퍼링크가 나타나거나 나타나지 않는 두 가지 형이다. 이 하이퍼링크들은 링크의 관련성이나 질에 의하여 모듈화 될 수 없다. 그러나 어떤 형태는 링크 모듈화에 매우 유용하다. 사용자는 네트워크를 탐색할 수 있도록 하기 위해 주어진 하이퍼텍스트 페이지로부터 링크 관계 강도를 할당할 필요가 있다. 그렇지 않으면 모든 연결은 동일한 연관성을 갖는 것으로 간주되며 사용자는 네트워크 내에서의 특정 위치를 선택적으로 탐색을 할 수 없게 된다. 이 모델에서 의도하는 대로 하이퍼텍스트 네트워크의 자동 재구성이 가능하도록 하기 위하여 하이퍼링크에 연결 가중치가 할당되도록 하였다. 이것은 실제로 존재하는 연결을 생성이나 삭제하지 않고 모듈화가 가능하도록 하며 연관성의 평가를 위하여 가중치를 사용한다.

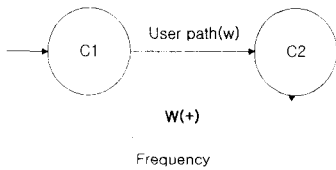
3.2 학습 규칙 모델

이 연구에서 제안하는 학습 규칙 모델은 다음의 세 가지이다. 이 규칙들은 사용자가 특정 연결을 선택하는지 여부에 따라 연결 강도를 부분적으로 변경할 수 있다.

3.2.1 사용빈도

사용빈도(frequency) 학습 규칙의 기능은 만약 두 개의 페이지가 시간적으로 연이어서 활성화되면, 두 페이지간의 연결 강도는 강화되어야 한다는 규칙이다. 이 강화 원칙은 자동화된 학습을 위한 많은 시스템과 모델의 핵심이다. 사용빈도 학습 규칙은 방문된 두 개의 하이퍼텍스트 페이지 사이의 강도를 강화하므로 주어진 연결이 사용되면 사용될수록 연결 강도가 높아진다.

if user visit c2 after c1 then
assign $W(c1,c2)$ to a value



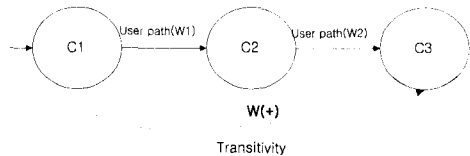
[그림 2] 사용빈도 학습규칙
[Fig. 2] Frequency learning rule

3.2.2 이행성

이행성(transitivity) 학습 규칙은 방문된 연결을 실제로 강화하지 않고 네트워크에서 새로운 연결을 생성한다. 사용자가 개념 c1 이라는 노드에서 개념 c2라는 노드로 향해하고 연속적으로 c2에서 c3라는 노드로 탐색할 때 이행성 학습 규칙은 c1과 c3 노드 사이의 연결을 강화한다. 그러므로 이행성 학습 규칙은 관련 있는 노드 사이에 링크를 놓음으로서 검색 경로를 짧게 만들어 효율성을 높이려는 것이다. 그러나 이행성이 새로운 연결을 도입할 때마다 이 연결은 만약 사용자가 새로운 연결이 유용하다고 판단하여 사용할 때만 성공한다고 할 수 있다. 다시 말해 이 규칙은 새로운 연결의 사용에 의한 사용빈도 학습 규칙을 적용 받게 되는 것이다. 그러나 사용자가 새로운 연결이 연관성이 없고 이 연결을 선택하지 않으면 이행성 학습 규칙에 의해 할당된 조기의 작은 상태로 남아있게 된다.

if user visit consequently c1, c2 and c3 then
{

add link c1 to c3
assign $W(c1,c3)$ to a value }

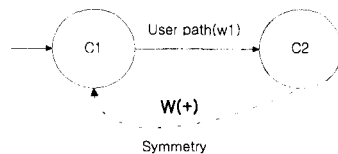


[그림 3] 이행성 학습 규칙
[Fig. 3] Transitivity learning rule

3.2.3 대칭성 학습 규칙

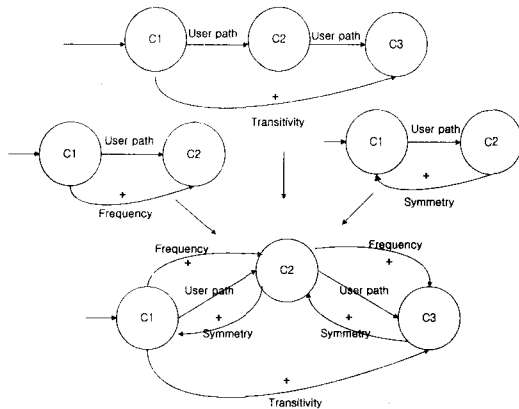
하이퍼텍스트 네트워크의 연결들은 방향성이 있으며 따라서 대칭적인 필요가 없다. 물론 어떤 하이퍼텍스트 페이지 c1는 다른 하이퍼텍스트 페이지 c2를 참조하고 그 반대로도 참조하지만 위의 경우는 아니다. 그럼에도 불구하고, 하이퍼텍스트 네트워크는 연상 네트워크이기 때문에 대칭성 형태를 재 생성하는 것이 타당할 것이다. 대칭성(symmetry) 학습 규칙은 C1 노드와 C2 노드 사이의 연결이 방문될 때마다 C2 노드로부터 C1 노드로의 모든 연결을 강화한다. 대칭성에 의하여 한번 강화된 연결은 사용자가 탐색 시 새로 형성된 연결을 방문하거나 방문하지 않을 수 있다. 만약 방문자가 그 연결을 선택한다면 사용빈도에 의하여 강화된다.

if user visit c2 after c1 then
assign $W(c2, c1)$ to a value



[그림 4] 대칭성 학습 규칙
[Fig. 4] Symmetry learning rule

이 세 가지 학습 규칙은 브라우징 동안 병렬적으로 엄격히 부분적으로만 동작한다.



[그림 5] 학습 규칙 기능의 전체 모형도

[Fig. 5] Schematic overview of learning rules function

3.3 링크 정렬

네트워크 구조를 재구성함으로써 링크의 연결 강도를 내림차순으로 정렬하여 사용자에게 제시할 수 있다. 네트워크 내의 어떤 페이지도 가장 높은 연결 강도의 링크가 맨 위에 나타난다. 선택 항목을 정렬하여 제시하는 원리는 많은 장점을 가진다. 최소한 축적된 탐색 습관에 대한 한 번의 철저한 분석으로 하이퍼텍스트 링크의 정렬된 순서와 선택될 확률 사이의 강한 관련성을 찾을 수 있다. 만약 시스템 기능이 의도한 대로라고 가정한다면, 주어진 어떤 페이지로부터의 적절한 연결은 가장 강한 연결 강도를 갖게 될 것이다. 그러므로 가장 강한 연결 강도를 갖은 링크가 선택될 확률이 가장 높기에 링크 목록의 상단에 나타난다. 이러한 링크 순서화는 선택 시간을 개선하고 인지적인 부하를 줄일 수 있어 효율적이다. 실세계의 하이퍼텍스트 시스템 연결 강도는 링크 컬러링, 폰트 크기 등 많은 다른 방법으로 보여질 수도 있지만 링크 순서화는 보다 더 효율적인 중요한 방법이다.

4. 실험 및 평가

이 장에서는 하이퍼텍스트 네트워크의 동적 재구성을 위한 학습 규칙을 포함한 제한된 프로토타입 시스템을 개발하고, 대학의 학생들을 대상으로 두 가지의 실험을 하고 그 결과를 평가한다.

4.1 실험

이 연구는 시스템의 효율성을 증진시키기 위한 시스템 개발보다는 하이퍼텍스트 네트워크를 재구성하기 위해 제안한 학습 규칙이 사용자의 메타 모델과 어떤 연관 관계를 갖고 있는지를 알아내는 것에 주안점을 두었기 때문에, 시스템은 실험을 위한 일반적인 웹 기반 프로토타입 시스템으로 구현하였다.

웹 상의 동적 하이퍼텍스트 시스템에서 브라우징 행위를 저장하여 이용하는 방법의 효과를 연구하기 위해 실험을 한다. 하이퍼텍스트 노드를 줄인 형태인 단어나 하이퍼링크만으로 구성하여 구현하였으며 특정 브라우징 전략을 실험 참석자들에게 설명하였다. 참석자들은 2개 대학의 학생들에게 홈페이지를 통해 모집하였으며, 실험에 참여한 2명에게 추첨을 통해 상품을 주는 것으로 하였다. 실험적 네트워크를 위해 100개의 자주 사용되는 명사 단어를 추출하여 노드로 사용하였다. 이 명사들은 구체적인 단어들에서부터 추상적 단어들까지 포함되어 있으며 기억이라는 단어와 연관된 단어들이 연관 정도에 따라 네트워크를 구성하는 것이다.

실험은 두 가지로 수행이 되었는데, 전체적으로 600명이 참석하였으며, 일주일 정도에 걸쳐 진행되었다. 서버 로그 파일을 보면 참석자들은 세션 당 10개의 링크를 선택하였다. 실험적 하이퍼텍스트 시스템은 분석을 위해 40개의 요청 당 하나의 네트워크 구조 복사본을 저장하였다.

애플리케이션은 네트워크 내에 모든 다른 노드에 연결에 가중치가 있고 시스템의 학습 규칙이 적용된 네트워크 노드를 포함하고 있다. 실험의 시작 시점에서, 애플리케이션은 하이퍼링크의 첫 번째로 순서화하기 위해 0.1 이하의 작은 난수 값으로 네트워크 안에 모든 단어 사이에 연결을 초기화한다. 프로그램을 처음 접속하면 각 참석자에게 할당된 네트워크에 시작 위치가 난수로 할당된다.

프로그램은 학습 규칙이 적용되었는데, 각 학습 규칙은 현재의 가중치에 독립적으로 서로 다른 절대값의 가중치를 두었다. 사용빈도는 연결 가중치는 1씩 증가, 이행성은 0.5를 증가한다. 이행성 학습 규칙의 증가 값이 낮은 것은 이 규칙이 사용자에게 의해 실제로 선택되어지지 않은 연결에 증가 값을 줄 수도 있으며, 이것은 사용빈도 학습 규칙에 반대되는

것이기 때문이다. 비록 이행성 연결이 많은 경우에서 유용하지만, 그렇지 않을 수도 있다. 대칭성 학습 규칙은 두 번째 실험에서만 적용이 되었으며, 0.3의 증가 값을 적용한다. 이 학습 규칙의 값은 이행성 규칙과 비교했을 때의 추정치를 반영한 것이다.

4.2 평가

실험은 임의의 구조로부터 시작하였지만, 모든 노드로부터 의미 있는 연결이 빠르게 진행되었다. 사용자가 네트워크에 임의의 위치로부터 브라우징을 시작하도록 되어있지만 브라우징 행위는 오히려 네트워크의 제한된 부분에 집중이 된 것처럼 보인다. 실험에 대부분의 노드는 타당한 빈도로 검색되고 브라우징 되었다. 800개 정도의 선택이 있는 후에는 생각, 아이디어, 연구 등과 같은 의미 있는 연결들이 리스트의 상단에 나타났으며, 두 가지의 실험 모두 800개 연결까지 매우 빠르게 진행되고 다음부터 천천히 느려지며, 접근선을 이루었다. 그러나 그 다음 순서로 나타난 법, 빛 등 처음 임의의 값을 배정한 초기 값들이 그대로 남아있어서 나타난 것들이었다. 1,200개 링크 선택이 있는 후에, 모든 순서화된 링크는 기억에 충분히 연관되었다고 보여지나 이 리스트의 링크 순서가 아직 정제되지 않았다. 1,800개의 선택 후에 새로운 연결이 나타나지 않을 만큼 안정적으로 연결이 구성되었다. 단지 링크의 순서가 약간씩 변할 뿐이었다. 기억으로부터 링크의 이러한 리스트는 기억의 의미적 또는 연관적 강도를 잘 반영한 위치에 연결이 구성되어 있는 안정적인 구조를 이루었다고 볼 수 있다.

학습 규칙들이 얼마만큼의 영향을 미치는지를 분석하기 위해, 두 번째 실험으로부터 네트워크 내 20개의 가장 높은 연결에 대한 학습 규칙에 의한 가중치들을 조사하였다. 20개 연결 중 처음 6개는 대칭성과 이행성 학습 규칙에 의한 가중치였고, 다음이 사용빈도 규칙에 의한 가중치가 반복적으로 나타났다. 노드 지식과 연구 사이에 연결을 위한 가중치와 노드 생명과 자연 사이에 연결 가중치를 네트워크의 링크의 수에 대한 함수로서 분석한 결과, 두 연결 모두 대칭성과 이행성 학습 규칙에 의해 주도되었으며, 이 결과로 이 연구에서 제안한 이행성과 대칭성이 의미 있는 역할을 한다는 것을 알 수 있다.

5. 결론

이 연구에서는 웹이 정보의 저장과 검색에 핵심적 매체가 된 만큼 하이퍼텍스트 네트워크에서 탐색과 검색의 효율성을 증진하기 위해 하이퍼텍스트 네트워크를 링크에 대한 사용자의 선호 패턴 그룹의 표현을 축적한 사용자 모델로 구조화하였다. 이것은 협력 필터링 기술의 장점과 탐색 지원 접근 방식을 혼합한 것이다. 이를 위해, 사용자의 브라우징 행위에 따라 자동적으로 하이퍼텍스트 네트워크를 재구성하는 사용빈도, 이행성, 대칭성 등의 세 가지의 학습 규칙 모델을 제안하였고, 실험을 위한 프로토타입 시스템을 개발하여 실험하였다. 시스템은 링크 가중치를 네트워크의 하이퍼링크에 부여하고 세 가지의 학습 규칙에 따라 링크 가중치를 변경할 수 있다. 네트워크 구조의 변경은 링크 가중치의 내림차순에 따라 동적으로 링크가 배열되어 사용자에게 제시됨으로써 동적 기능을 가질 수 있게 되었다. 학습 규칙 모델의 효율성을 보이기 위해 대학에서 지원자를 모집하여 일정기간 동안 실험을 하였다. 두 가지의 실험을 수행하였다. 첫 번째는 사용빈도와 이행성을 갖춘 것과 두 번째는 대칭성까지를 포함하여 실험하였다. 두 가지 실험에서 개발 속도를 비교한 것과 거의 동일한 시점에서 하이퍼텍스트 네트워크가 안정적인 형태를 자지고 있음을 알 수 있었다. 또한 두 개의 노드 사이에 세 개의 학습 규칙이 어떤 영향을 미치는가에 대한 분석에서, 이행성과 대칭성 학습 규칙이 사용빈도에 의미 있는 영향을 미치는 것을 알 수 있었으며, 이는 이 연구에서 제안한 모델의 효율성을 보여주는 것이라 할 수 있다.

향후 연구로는 보다 전체적으로 보다 많은 양의 실험을 통해 정확한 분석과 협력적 필터링 기법이 미치는 영향을 보다 상세히 분석하는 연구와 제안된 학습 규칙을 하이퍼텍스트 네트워크 설계를 위한 자동화된 도구에 적용하는 것이다.

※ 참고문헌

- [1] Alexa, <http://www.alexa.com>, 2001.

- [2] J. Bollen and F. Heylighen, "A System to Restructure Hypertext Networks into Valid User Models," *The New Review of Hypermedia and Multimedia*, 1998.
- [3] P. Brusilovsky, "Adaptive Hypermedia." *User Modeling and User-Adapted Interaction*, 11, 2001.
- [4] J. F. Cove and B. C. Walsh, "On-Line text Retrieval via Browsing," *Information Processing and Management*, 24(1), 1988.
- [5] P. DeBra and L. Calvi. "AHA! An Open Adaptive Hypermedia Architecture," *The New Review of Hypermedia and Multimedia*, 1998.
- [6] I. Im, *Augmenting Knowledge Reuse Using Collaborative Filtering Systems*, Ph.D. Dissertation, University of Southern California, 2001.
- [7] T. Joachims, D. Freitag, and T. Mitchell, "Web-watcher: A Tour Guide for the World Wide Web," *Proceedings of IJCAI*, 1997.
- [8] D. H. Jonassen, "Effects of semantically structured Hypertext Knowledge Bases," In C. McKnight, A. Dillon and J. Richardson, eds., *Hypertext: A Psychological Perspective*, Ellis Horwood, 1993.
- [9] C. Kaplan, J. Fenwick, and J. Chen, "Adaptive Hypertext Navigation Based on User Goals and Context," *User Models and User Adapted Interaction*, 3(2), 1993.
- [10] A. Kobsa, "Generic User Modeling Systems," *User Models and User Adapted Interaction*, 11, 2001.
- [11] G. Marchionini, *Information Seeking in Electronic Environments*, Cambridge University Press, 1995.
- [12] J. Nielsen, "The Art of Navigating through Hypertext," *Communications of the ACM*, 33(3), 1990.
- [13] C. Sherman and G. Price, *The Invisible Web*, Information Today, Inc., 2001.
- [14] R. Wilkinson. "User Modeling for Information Retrieval on the Web," *Proceedings of the 2nd Workshop on Adaptive Systems and User Modeling on the WWW*, 1998.

백 영 태



1989년 인하대학교
전자계산학과 졸업(이학사)
1993년 인하대학원
전자계산공학과 졸업
(공학석사)
2000년 인하대학원
전자계산공학과 박사수료
1998.5-1998.2 대상정보기술
(주) 정보통신연구소
선임연구원
2001.3 한국멀티미디어기술사
1998.3 - 현재 김포대학
컴퓨터계열 조교수
관심분야 : 멀티미디어/
하이퍼미디어시스템,
지능형 시스템,
웹 기반 교육 시스템

이 세 훈



1985년 인하대학교
전자계산학과 졸업(이학사)
1987년 인하대학원
전자계산학과 졸업(이학석사)
1996년 인하대학원
전자계산공학과 졸업
(공학박사)
1987-1990 해병대 전산실
분석장교
1991-1993 비트컴퓨터
기술연구소 선임연구원
1999.6 한국멀티미디어기술사
2000-현재 (주)중앙네트워크
시스템 연구소장
1993-현재 인하공업전문대학
전자계산기과 부교수
2001-현재 Visiting Scholar,
Dept. of Information
System, New Jersey Institute
of Technology, USA
관심분야 : 분산개체컴퓨팅,
소프트웨어공학, 원격교육,
멀티미디어/하이퍼미디어
시스템, XML/JAVA