

웹 미디어 시스템을 위한 확률 분포 함수와 사용자 프로파일에 기반 한 동적 적응 모델

백영태† · 이세훈† †

요 약

이 논문에서는 웹 미디어(웹 기반의 하이퍼미디어) 시스템에서 이산 확률 분포 함수와 사용자 프로파일 기반의 동적 적응 모델을 제안하였다. 이 모델은 응용 영역을 동적 적응 객체의 가중치 방향성 그래프로 표현하며, 사용자 행위를 이산 확률 함수를 동적으로 구축하는 접근 방식을 이용하여 모델링한다. 제한한 확률적 해석은 웹 미디어 구조에서 사용자의 탐색 행위를 추적하여 사용자 행위에 대한 잠재적 속성을 나타내는데 사용될 수 있다. 이러한 접근 방식은 사용자에게 가장 알맞은 프로파일을 동적으로 할당할 수 있다.

Dynamic Adaptive Model based on Probabilistic Distribution Functions and User's Profile for Web Media Systems

Yeong-Tae Baek† · Se-Hoon Lee† †

ABSTRACT

In this paper we proposed dynamic adaptive model based on discrete probabilistic distribution functions and user's profile for web media systems(web based hypermedia systems). The model represented that the application domain is modelled using a weighted direct graph and the user's behaviour is modelled using a probabilistic approach that dynamically constructs a discrete probability distribution functions. The proposed probabilistic interpretation of the web media structure is used to characterize latent properties of the user's behaviour, which can be captured by tracking user's browsing activity. Using that distribution the system attempts to assign the user to the best profile that fits user's expectations.

1. 서 론

웹을 미디어로 하는 정보 시스템은 정보의 사용자 중심 접근을 위한 새로운 컴퓨팅 환경으로 가장 대중성을 띄고 있으며, 이 논문에서는 이를 웹 미디어 시스템이라 정의한다. 이러한 웹 미디어 시스템은 전통적인 하이퍼미디어 시스템이 웹

을 기반으로 진화되어 새로운 정보 서비스 환경으로 자리 잡은 것이라 할 수 있다[4].

웹 미디어 시스템에서 프리젠테이션과 콘텐츠의 사용자의 요구와 목적에 따른 동적 적응 기능인 개인화는 중요한 요구 사항이 되고 있다[4,5]. 콘텐츠 개인화의 응용 분야는 온라인 광고, 웹 마케팅, 전자상거래, 온라인 교수 및 학습 등 다양하다. 이러한 시스템은 전형적으로 대규모 하이퍼공간을 사용자가 자유롭게 탐색할 수 있도록 하고 있다. 그러나 이러한 웹 미디어의 풍부한 링크 구조는 일부 심각한 유용성 문제를 발생

† 정 회 원 : 김포대학 컴퓨터계열 조교수
 †† 정 회 원 : 인하공업전문대학 컴퓨터정보공학부 교수
 논문접수 : 2002년 12월 22일, 심사완료: 2003년 1월 27일
 * 이 논문은 2002학년도 김포대학의 연구비 지원에 의하여 연구되었음

시키고 있다[1,6].

일반적인 웹 미디어 시스템은 모든 사용자에게 페이지 내에 같은 링크를 제시한다. 탐색 문제를 해결하기 위해, 시스템은 각 사용자별로 개인화된 링크나 사이트 맵이나 콘텐츠 목차 같은 탐색 도구들을 제공해야 한다. 따라서 시스템은 사용자의 관심, 선호도, 전문성 등의 정보를 바탕으로 가장 알맞은 도구들을 제시해야 한다.

웹 콘텐츠 개발자가 미처 예측하지 못한 탐색은 사용자에게 대한 이해력 문제를 발생시키며, 개발자는 모든 페이지에 대해 사용자가 해당 페이지를 액세스할 때 사용자에게 대해 가정(assumption)을 예측해 저작해야 한다. 그러나 모든 사용자에게 가장 알맞은 링크 구조를 모두 예측하고 개발한다는 것은 거의 불가능하다. 따라서 페이지는 항상 같은 방식으로 제시되며, 이것은 사용자가 이미 방문한 페이지일 경우 중복된 정보를 포함하게 될 것이고, 사용자에게 대한 부족한 예측으로 사용자는 원하는 정보를 찾기 위해 많은 시간을 보내야 할 것이다.

이러한 문제를 해결하고 보다 정교한 서비스를 위해 사용자 모델링과 사용자 인터페이스 기술을 결합한 개인화 기술에 대한 연구가 근래 시작되었다[2]. 근래 개인화 시스템은 사용자의 등록 정보의 이용뿐만 아니라 사용자 탐색 행위를 관찰해 자동적으로 사용자 프로파일(profile)을 변경하여 동적인 적응성을 갖고 있는 동적 적응 시스템(adaptive systems)이다.

동적 적응은 사용자 모델을 참조하여 영역 모델로부터 적절한 콘텐츠와 탐색 지원을 제공하여 개인화 목적을 달성하기 위한 기술로서, 주로 규칙을 기반으로 한 생성 시스템(production system)과 확률과 베이지안 기법 등의 통계적 기법을 사용한 평가에 대한 연구들이 있다[7].

이 논문에서는 웹 미디어 시스템에서 동적으로 개인화된 서비스를 할 수 있는 이산 확률 기법에 기반한 동적 적응 기법을 제안한다. 웹 미디어 시스템의 논리적 구조와 콘텐츠는 3개의 논리적 레벨에서 서술되며 각 레벨에서 동적 적응이 시도된다. 이산 확률 분포 함수는 각 프로파일에 속한 사용자가 탐색하는 동안 수정되는 확률 등이며, 이러한 분류를 이용해 시스템은 사용자가

원하는 것을 정확히 판단하여 사용자 환경에 가장 알맞은 프로파일을 할당할 수 있다.

2. 사용자 모델링과 동적 적응 기법

동적 적응은 일반적으로 사용자의 지식과 관심에 따라 동적으로 제시할 콘텐츠와 프리젠테이션, 링크 등을 생성하는 것이다. 이 장에서는 사용자 모델링 기법과 적응적 기법에 대해 고찰한다.

2.1 사용자 모델링

사용자에 관한 데이터는 크게 사용자의 정적 데이터, 동적 데이터, 환경 데이터 등의 세 가지로 나눌 수 있다. 정적 데이터는 인구통계학적 데이터, 사용자의 지식, 사용자 능력, 관심과 선호, 목적과 목표 등이며, 동적 데이터는 사용자의 행위를 실시간 모니터링 함으로써 얻는 데이터이다. 환경 데이터는 사용자가 갖고 있는 하드웨어와 소프트웨어, 환경적 특성 등에 관한 데이터이다. 이러한 사용자에게 관한 데이터를 획득하는 방법은 데이터의 성격에 따라 사용자로부터 직접 입력을 받거나 획득 규칙, 계획 인식, 스테레오 타입 추론 등 다양한 방법이 연구되고 있다[10].

사용자 모델링의 본질적 특성인 불확실성을 다루기 위해, 많은 시스템에서는 사용자 모델내에 사용자에게 관한 지식 정도를 숫자화하여 나타내고 있다. HYDRIVE는 항공기 수압 시스템에 고장 수리에 관한 학습자의 능력을 모델링하였으며, 베이시안 네트워크를 사용하였다[8]. 네트워크 노드의 확률 분포는 여러 추상적 레벨을 구성하는 “전자공학 지식”, “전략적인 지식” 같은 변수를 명확하게 나타내고 있다. 보다 구체적 레벨에서, 노드는 “해석된 행동”을 나타낸다. 그런 노드는 사용자 행동들의 관찰된 순서가 고정된 수의 행동 카테고리 중 하나에 속할 확률을 저장한다. 사용자 모델에 입력을 위해, HYDRIVE는 네트워크 노드를 생성하고 관찰된 행동 카테고리의 확률을 1로 하여 네트워크에 추가한다. 이것은 HYDRIVE의 획득 과정이다. 두 번째 추론은 해석된 행동 노드에서 “전략적 지식” 같은 보다 높

은 레벨 노드까지 확률을 전파하는 것에 의해 수행된다. 즉, 가능한 추론은 주로 베이지안 네트워크 구조와 링크에 연결된 조건적 확률에 표현된 미리 정의된 영역 지식에 의해 결정된다.

확률적 사용자 모델 표현에 사용된 다른 증거 기반 기법으로는 선형 파라미터, 퍼지 논리, Dempster-Shafer 이론(Dempster-Shafer theory) 등이 있다.

2.2 적응적 기법

웹 미디어 시스템은 서로 다른 사용자에게 내용 프리젠테이션과 탐색 지원을 이용하여 동적으로 적용할 수 있는 시스템을 말한다. 현존하는 웹 미디어 시스템은 교육용 하이퍼미디어 시스템, 온라인 정보 시스템, 온라인 도움 시스템, 정보 검색 하이퍼미디어 시스템, 공공 정보 시스템, 그리고 개인의 관심을 관리하는 시스템 등으로 구분할 수 있다[5].

동적 적응 웹 미디어에서 동적 적응 될 수 있는 것은 내용 단계와 링크 단계로 구분하고 첫 번째를 동적 적응 프리젠테이션, 두 번째를 동적 적응 탐색 지원이라고 부른다[4].

동적 적응 프리젠테이션은 현재 사용자의 지식, 목적 등 각 사용자들 간의 서로 다른 특징들을 가지고 사용자가 접근하려는 페이지의 내용을 동적 적응시키는 것이다.

동적 적응 프리젠테이션 방법으로는 추가 설명, 선행 설명, 비교 설명, 설명 방법의 변화, 재분류 등이 있다. 이 중에서 가장 호응이 좋은 방법은 추가적인 설명이 나오는 방법이다. 이 방법의 목적은 추가설명을 원하거나 필요로 하는 사용자들에게 추가적인 정보, 설명, 실례 등을 제공하는 것이다. 동적 적응 프리젠테이션에 사용되는 기술로는 조건 텍스트, 텍스트 펼치기, 프래그먼트 다양화, 페이지 다양화, 프레임기반 기술 등이 있다.

동적 적응 탐색 지원은 사용자의 목적과 지식, 그리고 개별적인 사용자의 다른 특징들을 가지고 프리젠테이션되는 링크의 방법을 동적 적용함으로써 하이퍼공간 내의 경로를 사용자가 찾을 수 있도록 도와주는 것이다.

동적 적응 탐색 지원에서 사용되는 방법으로 전역적 안내, 지역적 안내, 방향 지원 등이 있다. 전역적 안내는 사용자가 한 주제에 대해 학습하고 싶을 때 사용자가 원하는 것을 명확히 지적해주는 것과 사용자가 의미 있는 페이지들을 읽도록 순서를 부여하여 사용자에게 제안하여 다양한 범위에서 시스템이 탐색 경로를 제안하는 것으로 교육용 하이퍼미디어에서 유용하다.

동적 적응 탐색 지원에서 사용되는 기술로는 링크 숨기기, 링크 주석, 직접 안내, 링크 분류, 사이트 맵 등이 있다. 현재 가장 많이 사용되는 동적 적응 탐색 지원 기법은 링크 숨기기이다. 기본 원리는 학습자의 현재 목적과 적절한 관련이 없거나 읽을 필요가 없는 링크를 숨김으로써 탐색 공간을 제한하기 위한 것이다. 링크 분류는 사용자모델에 표기된 학습자 특성에 따라 학습자에게 적합한 노드를 가장 쉽게 인식할 수 있는 방법을 기준으로 특정 페이지의 모든 링크를 분류하여 제시하는 방법이다[3].

3. 확률적 해석기반의 동적 적응 모델

이 장에서는 웹 미디어에서 응용영역을 모델링하기 위한 스키마구조와 사용자를 동적으로 분류하기 위한 모델을 제안한다.

3.1 응용 영역 스키마 구조

이 절에서는 웹 미디어 스키마 구조에서 아크의 가중치에 대한 확률적 해석을 제안한다. 이것은 웹 미디어 구조에서 사용자의 탐색 행위를 추적하여 사용자 행위에 대한 잠재적 속성을 바탕으로 사용자 분류를 하는데 사용된다. 즉, 사용자 모델은 초기에 스테레오타입(stereotype)으로 구축하고, 실행 시에 사용자 행위를 모니터링해서 사용자의 행위에 따라 사용자 모델의 프로파일을 동적으로 조정하는 것이다.

이 논문에서는 전체 응용 영역을 동적 적응 콘텐츠 객체의 가중치 방향성 그래프로 생각한다. 이를 수식으로 나타내면, M개의 다른 프로파일을 갖는 응용 영역과 동적 적응 콘텐츠 객체의 집합을 N이라 하면, 프로파일 $k=1, \dots, M$ 에 대해 i

∈N인 i라는 노드에 j라는 도착 노드까지의 출력 링크를 집합 L_{ik} 라 하자. 가중치 방향성 그래프 $G=(N,E)$ 에서 N은 각 노드를 나타내고 E는 각 방향성 아크를 나타내며, 이를 수식으로 표현하면 다음과 같다.

$$G_k=(N_k, E_k)$$

$$E=L_{i1} \cup L_{i2} \cup \dots \cup L_{iM} = \bigcup_{k=1, \dots, M} L_{ik}$$

수식을 단순화하기 위해 가중치 방향성 그래프 G에서 각 프로파일에 대응하는 노드와 아크를 가중치 그래프 $G_k(k=1, \dots, M)$ 라하고, 이 G_k 를 논리적 탐색 그래프라 한다.

그러면 G_k, N_k, E_k 는 다음과 같이 정의할 수 있다.

$$G_k=(N_k, E_k)$$

$$N_k=\{i \mid (i, j, k) \in E \vee (j, i, k) \in E\}$$

$$E_k=\{(i, j) \mid (i, j, k) \in E\}$$

E_k 에서 (i,j) 아크의 가중치를 $W_k(i, j)$ 라 하고 이것을 조건부 확률 $P(k \mid i, j)$ 로 정의하자. 즉, 프로파일 k에 속한 사용자가 i 노드에 도착해서 j 노드에 링크될 조건부 확률로 다음과 같이 정의된다.

$$W_k(i, j): E_k \rightarrow [0, 1]$$

$$W_k(i, j) = P(j \mid k, i), (i, j) \in E_k, k=1, \dots, M$$

$P(k \mid i, j)$ 는 한 노드에서 그 자신으로 링크가 될 확률로, 불가능한 경우를 나타내며 따라서 확률은 영(0)이고, 각 노드 i에서 각 프로파일로의 출력 아크의 조건부 확률의 합은 항상 1이 된다. 즉,

$$\forall i \in N_k, W_k(i, 1) + W_k(i, 2) + \dots + W_k(i, M)$$

$$= \sum_{j=1}^M W_k(i, j) = 1$$

이고, 조건부 확률로 보면

$$P(1 \mid k, i) + P(2 \mid k, i) + \dots + P(M \mid k, i) = 1$$

이 된다.

그래프 G_k 에서 패스 S를 순서 있는 노드의 집합으로 다음과 같이 정의한다.

$$S=\{S_0, S_1, \dots, S_l \mid (S_j, S_{j+1}) \in E_k, j=0, \dots, l-1\}$$

조건 $(S_j, S_{j+1}) \in E_k$ 가 다른 논리적 탐색 그래

프를 포함하는 패스를 허락하도록 완화하기 위하여 패스에 대한 표준적 아크 기반 정의를 사용하지 않는다. 이것은 프로파일 k를 갖는 사용자가 노드 S_j 에서 S_{j+1} 로의 링크를 선택하고 사용자가 새로운 프로파일 h로 이동된다면, G에 속하는 것으로 판단하여 다른 조건 $(S_{j+1}, S_{j+2}, h) \in E$ 을 고려하게 된다.

논리적 탐색 그래프에 포함되는 패스는 사용자 탐색 행위 동안 다른 프로파일로 이동된 사용자의 경우에도 패스가 허락된다.

프로파일 k에 속한 사용자가 패스 S를 따를 확률은 다음과 같다.

$$\begin{aligned} P_s^k &= W_k(S_0, S_1) \times \dots \times W_k(S_{l-1}, S_l) \\ &= \prod_{j=0, \dots, l-1} W_k(S_j, S_{j+1}) \\ &= P_s^k(S_1 \mid K, S_0) \times P_s^k(S_2 \mid K, S_1) \times \dots \times P_s^k(S_l \mid K, S_{l-1}) \end{aligned}$$

여기서 S_{ij}^k 는 프로파일 k에 속하는 아크를 통하는 노드 i와 j 사이의 일반적인 패스이다. P_s^k 는 패스 S에 속하는 아크에 확률을 계산하는 것으로, 주어진 프로파일 k에 대해 i에서 j의 노드 사이의 패스 중 가장 짧은 패스를 S_{ij}^k 라 하고 프로파일 k에 대해 i에서 j노드 사이에 패스 S의 확률 중 최대 확률을 P_{ij}^k 라 하자. 그러면 P_{ij}^k 는 다음과 같이 정의할 수 있다.

$$\bar{P}_{ij}^k = \max_{S_{ij}^k} (P_{S_{ij}^k}^k)$$

이 논문에서는 웹 미디어 구조의 본질적 속성과 3개의 이산 확률 분포 함수(PDF; discrete Probability Distribution Function)로 모델을 제안한다.

1. $\mu(k)$ 는 각 프로파일 k에 대해, G_k 에서 가장 짧은 패스의 확률의 평균값에 비례하는 값으로 다음과 같이 정의된다.

$$\mu(k) = \frac{\sum_{q=1}^M \left[\sum_{(i, j) \in E_k} P_{ij}^k \delta(k-q) \right]}{\sum_{q=1}^M \sum_{(i, j) \in E_k} P_{ij}^k} \quad (\text{식 1})$$

이 이산 확률 분포 함수의 값이 높은 것은 웹 미디어 내에 가장 알맞은 패스가 있다는 것을 의미한다.

2. $p(k)$ 는 각 프로파일 k에 대해, G_k 에서 가장

짧은 패스의 길이의 평균값에 비례하는 값으로 다음과 같이 정의하자.

$$p(k) = \frac{\sum_{q=1}^M \left[\sum_{(i,j) \in E_c} |S_{ij}^q| \delta(k-q) \right]}{\sum_{q=1}^M \sum_{(i,j) \in E_c} S_{ij}^q} \quad (\text{식 2})$$

이 이산 확률 분포 함수 값이 높다는 것은 전체적인 개별화 과정에 장점이 될 수 있는, 웹 미디어 내에 보다 긴 알맞은 패스를 의미한다.

3. $n(k)$ 는 각 프로파일 k 에 대해, 해당 프로파일에 속한 노드의 수로 정의한다.

$$n(k) = \frac{\sum_{q=1}^M [|N_q| \delta(k-q)]}{\sum_{q=1}^M |N_q|} \quad (\text{식 3})$$

이들 값은 시간에 따라 변한다는 것을 유념해야 한다. 웹 미디어 구조는 노드나 아크, 가중치 등이 추가되거나 제거될 수 있는 등 동적으로 수정될 수 있다. 이는 여러 사용자의 행위를 일부 자동화된 관찰에 근거하거나 저작자에 의해 응용 영역의 증가된 지식에 근거하여 수정된다.

프로파일의 본질적 관련성을 나타내는 가중치 중앙값(weighted medium)은 다음과 같이 계산된다.

$$s(k) = \frac{\beta_0 \mu(k) + \beta_1 n(k) + \beta_2 p(k)}{\beta_0 + \beta_1 + \beta_2} \quad (\text{식 4})$$

여기서, $\mu(k)$ 와 $p(k)$ 의 값은 노드가 없는 프로파일은 높은 확률을 갖는 패스를 가질 수 없기 때문에 서로 적절하게 조절되어야 한다. $s(k)$ 에서 각 항의 높은 값은 프로파일 k 에 대한 높은 관련성을 나타내며, 따라서 $\beta_i > 0$ 이다.

3.2 사용자 동적 분류

이 논문에서 제안한 확률적 해석은 웹 미디어 구조에서 사용자의 탐색 행위를 추적하여 사용자 행위에 대한 잠재적 속성을 나타내는데 사용될 수 있다. 동적 적응 차원에서 사용자 행위에 관련된 속성은 사용자에 관한 스테레오타입 모델에 연관해서 나타낼 수 있다.

제안된 시스템은 이산 확률 분포 함수 $A(k)(k=1, \dots, M)$ 을 구축하는데, 각 그룹에 사용자의 "확률적 소속" 즉, 얼마나 많이 각 프로파일이 사

용자에게 맞는지를 측정하는 것이다. 즉, 사용자 행위에 기반해서 시스템은 동적으로 사용자에게 가장 적합한 프로파일을 할당한다.

탐색은 시작 노드에 연결된 최종 페이지부터 시작한다. 만약 사용자가 이미 등록되어 있다면, 마지막 $A(k)$ 가 현재로 지정된다. 아니면, 사용자는 일반적 프로파일을 할당받거나 질문을 기반으로 해서 계산된 프로파일을 할당받을 수 있다. $A(k)$ 의 초기값은 $A_0(k)$ 라고 한다.

노드 R_{r-1} 를 방문한 사용자가 다음 링크를 요구했을 때, 시스템은 새로운 이산 확률 분포 함수 $A'(k)$ 를 사용자 행위 변수와 가중치 중앙값 $s(k)$ 를 기초로 계산하고, 사용자에게 할당될 새로운 프로파일을 결정한다.

프로파일을 매번 변경하는 것을 피하기 위해, 고정된 일정 시간사이에 $A'(k)$ 분산을 평가하여 일정한 기간동안 프로파일을 유지하는 것도 가능하다. 프로파일의 유지 시간은 탐색 링크의 수 등과 같은 요소를 통해 계산할 수 있다.

사용자 행위는 사용자 행위 변수의 집합으로 저장된다.

현재의 프로파일 K_c

- 현재 이산 확률 분포 함수 $A(k)(k=1, \dots, M)$ 는 각 프로파일에 사용자의 소속 확률을 측정한다.
- 최근 탐색한 패스 $R = \{R_1, \dots, R_{r-1}, R_r\}$ 는 최근 방문한 노드를 포함한다. R_{r-1} 은 현재 노드이고 R_r 은 다음 노드이다. 최근 아크 (R_{r-1}, R_r, K_c) 는 사용자에 의해 선택된 출력 링크이다.
- 최근 노드에서 시간소모는 $(R_1), \dots, (R_{r-1})$ 이다.

$P_{R_r}^*$ 는 프로파일 k 에 속한 아크를 통해 패스 R 을 따르는 확률로 정의하고 $P_{R_r}^*$ 의 높은 값은 R 에서 방문된 노드가 프로파일 k 에 관련된다는 것을 나타낸다. 이것은 실제 패스가 프로파일 k 에 대해 알맞다는 것을 나타낸다.

$\overline{P}_{R_1, R_r}^*$ 는 프로파일 k 에 속한 아크를 통해 처음 노드 R_1 으로부터 시작하여 다음 노드 R_r 으로의

도착율로 정의하고 이 항목은 사용자가 다음 노드 R_r 에 도착할 수 있는 방법을 고려한다. 여기서 값이 높다는 것은 프로파일 k 의 링크를 통해 매우 자연스럽게 도착할 수 있는 길이 있다는 것을 의미한다.

$D[k]$ 는 노드 R_1 에서 R_r 로 방문한 노드에서 소비한 시간에 대한 가중치를 각 프로파일 k 에 대한 분포로 정의한다. 예를 들어, $\{n_1, n_2, n_3\}$ 가 최근에 방문한 노드이고 $\{t_1, t_2, t_3\}$ 가 각 노드에서 소비한 시간이라 하자. 노드 n_1 이 프로파일 k_1 과 k_2 에 속하고, 노드 n_2 가 k_2, k_3 에 속하고, 노드 n_3 가 k_1, k_4 에 속한다면, 분산은 다음과 같이 계산된다.

$$D[k] = [(k_1, t_1 + t_3), (k_2, t_1 + t_2), (k_3, t_2), (k_4, t_3)]$$

$D[k]$ 는 방문한 노드 상에서 얼마나 시간을 보냈는지를 보여주는데, 이것은 프로파일에 대한 분산이고 사용자의 관심을 나타내는 지수가 된다. 방문 시간은 정확해야 하며, 정확한 계산식에 관한 흥미로운 접근 방식이 [9]에서 제안되었다. 이것을 기초로 3개의 이산 확률 분포 함수를 정의한다.

1. $c(k)$ 는 프로파일 k 에 속하는 아크를 통해 R 패스를 따를 확률 P_R^k 로 다음과 같이 정의된다.

$$c(k) = \frac{\sum_{i=1}^M [P_R^i \delta(k-i)]}{\sum_{i=1}^M P_R^i} \quad (식 5)$$

2. $r(k)$ 는 처음 노드 R_1 로부터 다음 노드 R_r 로 프로파일 k 에 속한 아크를 통해 도달할 가능성으로 다음과 같이 정의된다.

$$r(k) = \frac{\sum_{i=1}^M [\overline{P}_{R_1, R_r}^i \delta(k-i)]}{\sum_{i=1}^M \overline{P}_{R_1, R_r}^i} \quad (식 6)$$

3. $t(k)$ 는 R_1 에서 R_{r-1} 로 방문된 노드의 소요한 시간을 갖는 가중치의 분포 $D[i]$ 로 다음과 같이 정의된다.

$$t(k) = \frac{\sum_{i=1}^M [D[i] \delta(k-i)]}{\sum_{i=1}^M D[i]} \quad (식 7)$$

사용자의 관심이 바뀌지 않는 임시적인 이탈은 $A(k)$ 상에서 $c(k)$ 와 $r(k)$ 의 영향을 적절히 조절하는 것이 고려될 수 있다. 전자는 실제 이동한 패스

를 고려하여, 프로파일을 최근 선호도에 따라 이동시키는데 목적이 있다.

반면 후자는 최근 선택을 무시하여, 방문한 노드 R_1 과 R_r 사이를 고려하지 않고 가장 짧은 패스에 대한 것이다. 마지막으로 프로파일의 동적 관련성을 나타내는 가중치 중앙값은 다음 식으로 계산된다.

$$d(k) = \frac{a_0 c(k) + a_1 r(k) + a_2 t(k)}{a_0 + a_1 + a_2} \quad (식 8)$$

$d(k)$ 에서 각 항목의 높은 값은 프로파일 k 에 대한 높은 관련성을 나타내며, 따라서 $a_i > 0$ 이다.

사용자가 속한 프로파일을 계산하는 주요 아이디어는 사용자의 동적 행위를 결합하는데 있다. 사용자의 동적 행위는 주로 위상 기하학에 따른 웹 미디어 스키마 $s(k)$ 의 구조적 속성과 통합된 $d(k)$ 이다. <알고리즘 1>은 사용자 행동에 기반한 새로운 확률 분포 함수 $A'(k)$ 를 계산한다.

<알고리즘 1> 새로운 확률 분포 함수 계산

<p>INPUT</p> <p>이산 확률 분포 함수 $A(k), A_d(k), s(k)$</p> <p>최근의 방문한 패스 $R = \{R_1, \dots, R_{r-1}, R_r\}$</p> <p>최근에 방문한 노드에서 소비 시간, $t(R_1), \dots, t(R_{r-1})$</p> <p>OUTPUT</p> <p>새로운 확률 분포 함수 $A'(k)$</p> <p>STEP</p> <p>새로운 이산 확률분포 함수 $d(k)$ 계산</p> <p>새로운 이산 확률분포 함수 $A(k)$ 계산</p> $A'(k) = \frac{\gamma_0 A_0(k) + \gamma_1 A(k) + \gamma_2 d(k) + \Delta \gamma_3 s(k)}{\gamma_0 + \gamma_1 + \gamma_2 + \Delta \gamma_3}$ <p>만약, $s(k)$가 변화하면 $\Delta=1$, 그렇지 않으면 $\Delta=0$이다.</p>

새로운 $A'(k)$ 는 4 가지 항목의 가중치 평균으로 계산되는데, 첫 번째 항목은 초기 사용자의 선택을 나타낸다. 두 번째 항목은 상호작용의 스토리이다. 세 번째 항목은 단일 사용자의 동적 행위를 나타내는 반면, 네 번째 항목은 웹 미디어의 구조적 속성을 나타낸다.

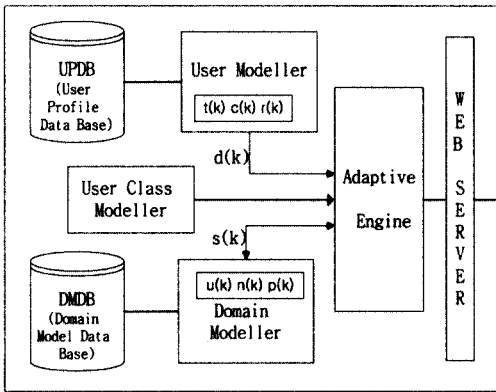
$A'(k)$ 에서 각 항목의 높은 값은 프로파일 k 에 대한 높은 관련성을 나타낸다. 따라서 $\gamma_i > 0$ 이 된다. 새로운 프로파일은 $A'(k)$ 분산에서 난수적 추출을 통해서 선택하거나 가장 높은 $A'(k)$ 값을 참고한다

4. 실험 및 평가

이 장에서는 제안된 모델을 기반으로 프로토타입 시스템을 설계하고 실험하여 모델의 타당성을 보인다.

4.1 동적 적응 시스템 설계

시스템은 (그림 1)과 같이 전체적으로 3개의 모델러와 2개의 DB로 구성된다. 영역 모델러는 웹 미디어의 정적 특성들을 모델링하며, 사용자 모델러는 동적 특성들을 모델링하게 된다. 동적 적응 엔진은 정적 특성과 동적 특성들을 이용하여, 사용자에게 가장 알맞은 최적의 프로파일을 동적으로 할당하게 된다. 시스템은 윈도우 2000 서버 환경에서 MS-SQL 2000, 웹 서버는 아파치 2.0, 웹 서버 스크립트 컨테이너는 JDK 1.4와 Tomcat 4.1, 문서의 동적 변환을 위한 COCOON 2.0을 이용하였다.



(그림 1) 동적 적응 시스템

다음은 시스템의 간단한 수행 절차이다.

1. 하이퍼미디어의 구조로부터 계산된 구조적 속성을 분석한다.
2. 행위를 시뮬레이션 할 필요가 있는 전형적인 사용자 그룹의 집합을 정의하는데, 사용자 클래스 모델러에 의해서 정의하며 이것은 시스템의 응답을 검증한다. 여러 다른 사용자 마스크를 각 클래스에 할당할 수 있고, 따라서 사용자의 행위는 시스템과 같은 상호 작용동안 변경될 수 있다. 사용자 클래스에 의해 모델링 된 행위는 임의의 방문 시간이나 아크의 선택을 포함한다.
3. 동적 적응 엔진에 의해 시뮬레이션이 실행

된다. 동적 적응 엔진은 멀티스레드 머신으로 여러 사용자의 요청을 생성하고 시각적인 방법으로 로그 결과를 보여준다.

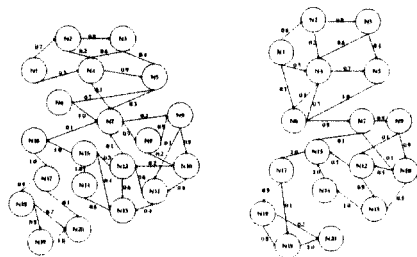
4. 사용자 모델러의 응답으로 사용자 클래스에 관한 프로파일 할당 결정을 분석한다.

5. 최종적으로, 확률 분포 함수 결과의 여러 가지 결과에 따라 다음을 실행한다.

첫째, 이 알고리즘에 사용된 파라미터를 조절한다. 즉, 불확실한 시간적 윈도우의 길이나 확률 분포 함수의 가중치에 사용된 파라미터의 값을 저작자가 입력한다. 둘째, 하이퍼미디어 구조를 조정한다. 이러한 작업들은 저작자의 경험에 따른 판단에 의존적이며, 여러 번의 시뮬레이션을 통해 가장 알맞은 값을 결정할 수 있다. 저작자의 결정을 도와줄 수 있는 시뮬레이터에 대한 연구가 필요하다.

4.2 프로그래밍 영역 스키마 구조와 동적 적응

프로그래밍 영역을 모델링하기 위해 먼저 스키마 구조를 정의하여야 한다.



1) 프로파일 1의 패스 2) 프로파일 2의 패스



3) 프로파일 3의 패스

(그림 2) 프로파일별 패스에 따른 아크의 가중치 확률

(그림 2)는 재채와 클래스의 개념에 대한 추상 개념 단계에서 각 프로파일별 링크에 가중치가

부가된 방향성 그래프로 구성한 형태이다. 각 프로파일은 자바 프로그래밍에서 초보자, 중급자, 전문가로 분류하고 학습 목적별로 분류하였다.

(그림 3) 한 노드에서 구성된 프로파일별 화면

(그림 3)은 객체와 클래스 개념 설명의 한 노드에서 구성된 프로파일별 화면이다. 하나의 노드는 동적 적응 콘텐츠 객체를 의미하고 하나의 동적 적응 콘텐츠 객체 즉, 하나의 페이지에는 여러 개의 링크가 프로파일별로 다르게 구성되어 있다.

<표 1> 프로파일 1에 대한 일반적인 패스와 확률 값

	일반적인 패스	확률(p_i)
S^1	{1,2,3,4,5,6,7}	0.21
S^2	{1,2,3,4,5,7}	0.09
S^3	{1,2,3,4,7}	0.03
S^4	{1,2,3,5,6,7}	0.16
S^5	{1,2,3,5,7}	0.07
S^6	{1,2,4,5,6,7}	0.09
S^7	{1,2,4,5,7}	0.04
S^8	{1,2,4,7}	0.01
S^9	{1,4,5,6,7}	0.19
S^{10}	{1,4,5,7}	0.08
S^{11}	{1,4,7}	0.03

모든 프로파일 중 프로파일 1에서 보면 시작 노드 N1에서 목적 노드 N7로 갈 수 있는 일반적인 패스와 패스 S의 확률 값은 <표 1>과 같다.

노드 1을 제외한, 시작 노드 Ni에서 목적 노드 N7로의 갈 수 있는 일반적인 패스와 패스 S에

따르는 확률 값은 <표 2>와 같다

<표 2> 패스 S에 따르는 프로파일 1의 일반적인 패스와 확률 값

N2-N7		N3-N7		N4-N7		N5-N7		N6-N7	
{2,3,4,5,6,7}	0.30	{3,4,5,6,7}	0.38	{4,5,6,7}	0.63	{5,6,7}	0.7	{6,7}	1.0
{2,3,4,5,7}	0.13	{3,4,5,7}	0.16	{4,5,7}	0.27	{5,7}	0.3		
{2,3,4,7}	0.05	{3,4,7}	0.06	{4,7}	0.1				
{2,3,5,6,7}	0.22	{3,5,6,7}	0.28						
{2,3,5,7}	0.10	{3,5,7}	0.12						
{2,4,5,6,7}	0.13								
{2,4,5,7}	0.05								
{2,4,7}	0.02								

프로파일 1에 대해 N1과 N7의 두 노드 사이에 가장 짧은 패스 \bar{S}_{17}^1 는 $S^{11}=\{1,4,7\}$ 이고 확률의 최대값을 가지는 패스 \bar{P}_{17}^1 는 $P^1_{S^1}=0.21$ 이다. 그 외 프로파일 1에 대해 두 노드 사이에 가장 짧은 패스 \bar{S}_{ij}^1 와 확률의 최대값을 가지는 패스 \bar{P}_{ij}^1 는 <표 3>과 같다.

<표 3> 두 노드 사이에 가장 짧은 패스와 확률의 최대값

\bar{S}_{ij}^1	i=1	i=2	i=3	i=4	i=5	i=6	$\sum \bar{S}_{ij}^1$
k=1	2	2	2	1	1	1	9
k=2	2	3	3	2	2	1	13
k=3	3	2	0	2	1	1	9

\bar{P}_{ij}^1	i=1	i=2	i=3	i=4	i=5	i=6	\bar{p}_i^1
k=1	0.21	0.30	0.38	0.63	0.70	1.00	3.22
k=2	0.19	0.30	0.38	0.63	0.90	0.90	3.30
k=3	0.36	0.45	0.00	0.56	0.80	1.00	3.17

(식 1), (식 2), (식 3)에 의하여 계산된 결과인 각 프로파일에 대한 하이퍼미디어의 구조적 속성 값은 <표 4>와 같다.

<표 4> 각 프로파일에 대한 하이퍼미디어의 구조적 속성값

프로파일	$u(k)$	$n(k)$	$p(k)$	$s(k)$
1	0.33	0.35	0.29	0.32
2	0.34	0.35	0.42	0.37
3	0.33	0.30	0.29	0.31

4.3 사용자 분류

현재 노드 N3를 방문한 사용자가 다음 링크로 노드N4가 올 경우 시스템은 새로운 이산 확률 분포 함수 $A'(k)$ 를 사용자 행위 변수와 가중치 평균 $s(k)$ 를 기초로 계산하고, 사용자에게 할당될

새로운 프로파일을 결정한다.

각 프로파일에 대해, 현재 이산 확률 분포 함수 $A(k)(k=1, \dots, M)$ 는 각 프로파일에 사용자의 확률로 저작자가 초기에 할당한 값으로

- $A(1) = 0.33,$
- $A(2) = 0.33,$
- $A(3) = 0.33$ 로 놓자.

각 노드에서 각 프로파일별 사용자가 소비한 시간은 <표 5>와 같다.

<표 5> 프로파일별 노드에서 소비한 시간 (단위 : 초)

프로파일 \ 노드	N1	N2	N3	N4
1	60	55	45	45
2	42	38	30	45
3	35	35		42

각 프로파일별 사용자가 최근 탐색한 패스 R 과 아크를 통해 처음 노드 R_1 으로부터 시작하여 다음 노드 R_2 로의 도착을 그리고 방문한 노드에서 소비한 시간에 대한 가중치는 <표 6>과 같다.

<표 6> 프로파일별 패스와 시간에 대한 속성값

프로파일	R	$\bar{P}_{R, R}$	T
1	{1,2,3,4}=0.34,	0.34	160
2	{1,2,3,4}=0.29,	0.30	110
3	{2,4}=0.80	0.64	70

사용자가 속한 프로파일을 계산하는 주요 아이디어는 사용자의 동적 행위를 결합하는데 있다. 단일 사용자의 동적 행위 값과 하이퍼미디어 구조적 속성값은 <표 7>과 같다.

<표 7> 단일 사용자의 동적 행위 값과 하이퍼미디어 구조적 속성값

프로파일	c(k)	r(k)	t(k)	d(k)	s(k)
1	0.24	0.27	0.47	0.29	0.35
2	0.20	0.23	0.32	0.24	0.35
3	0.56	0.50	0.21	0.47	0.30

<알고리즘 1>에 의해 사용자의 행동에 기반한 새로운 확률 분포 함수 $A'(k)$ 를 계산하면, $A'(1)=0.33, A'(2)=0.31, A'(3)=0.36$ 이 된다. 따라서, 사용자의 행동을 2가지로 분석할 수 있다.

첫 번째는 프로파일 1과 프로파일 2는 $A'(k)$ 의 값이 저작자가 설정한 초기값 0.33보다 차이가

10% 이내이므로 프로파일을 변경하지 않고 $A'(k)$ 이 $A(k)$ 로 적용하여 다음 노드로 진행하고, 두 번째 프로파일 3은 $A'(k)$ 의 값이 0.33보다 차이가 10% 이상이므로, 다음 노드 진행 시 프로파일을 변경하고 사용자에게 적합한 화면을 제공하여 동적으로 적용시킨다. 그리고 다음 노드 진행 시에는 $A'(3)$ 은 초기값으로 설정한다.

계속해서 노드 N4를 방문한 사용자가 다음 링크로 N5를 선택할 경우 <알고리즘 1>에 의해 새로운 확률 분포 함수 $A'(k)$ 를 계산하면, <표 8>과 같이 $A'(1)=0.34, A'(2)=0.33, A'(3)=0.32$ 이 된다. 따라서 노드 N5를 선택하였을 경우에는 각 프로파일별 확률 분포 함수가 0.03사이에 있으므로 다음 노드로 진행한다.

<표 8> 사용자 행동에 기반 한 새로운 확률 분포 함수 계산 값

프로파일	N3→N4		N4→N5		...
	속정값 $A'(k)$	적용된 $A(k)$	속정값 $A'(k)$	적용된 $A(k)$	
1	0.33	0.33	0.34	0.34	...
2	0.31	0.31	0.33	0.33	...
3	0.36	0.33	0.32	0.32	...

4.3 동적 적응의 실효성 분석 및 평가

개발한 시스템은 대학의 컴퓨터계열 1학년 학생 80명을 대상으로 실시하였으며, <표 9>와 같이 전체적으로 각각 20명씩 네 개의 그룹으로 분리하여 실험하였다.

<표 9> 실험을 위한 각 그룹별 동적 적응 기능 할당

구분	동적 적응 있음	동적 적응 없음
제어구조	그룹1	그룹2
객체지향 개념	그룹3	그룹4

실험은 자바 프로그래밍 언어에서 제어구조와 객체지향 개념의 2개의 장을 선정하였고 시스템을 이용해 동적 적응 기능이 있는 것과 동적 적응 기능이 없는 것을 개발하였다.

실험은 4주 동안 진행되었고, 처음 2시간 동안 학생들은 시스템의 사용법과 특징들에 대한 소개를 하였다. 소개 후 1시간 동안 시스템을 사용했고 질의응답 시간이 주어졌다. 질의응답 시간을

통해 대부분의 학생들이 사용에 어려움을 느끼지 않고 있음을 알 수 있었다.

<표 10> 그룹별 탐색 횟수의 통계량

	학생 수	최대 값	최소 값	평균	중간 값	표준 편차	분산
그룹1	20	63	51	55.80	56	2.78	7.75
그룹2	20	77	62	67.40	68	3.22	10.36
그룹3	20	67	48	52.11	51	4.80	23.04
그룹4	20	76	67	72.28	73	2.68	7.17

실험에서 각 그룹별로 학습목표를 얻을 때까지 탐색 한 횟수를 측정 한 결과를 요약하면 <표 10>과 같다.

동적 적용이 있는 그룹 1과 3이 동적 적용 없는 그룹2와 4보다 탐색 횟수가 상당히 낮은 결과를 얻었으며, 이는 제안 모델과 시스템이 학생들의 수준에 따라 동적 적용을 수행한다고 평가할 수 있다. 또한 실험의 영역별 즉, 제어 구조를 학습한 그룹 1, 그룹 2와 객체지향 개념을 학습한 그룹 3, 그룹 4를 비교해 보면, 객체 지향 개념을 학습한 그룹 3과 그룹 4가 보다 많이 동적 적용을 하고 있음을 알 수 있다. 이는 객체 지향 개념이 제어 구조보다 보다 복잡하고 어려운 점에서 원인을 찾을 수 있으며, 모델과 시스템이 보다 복잡한 응용에서 동적 적용이 많이 발생하는 것을 알 수 있다. 각 그룹의 탐색 횟수를 그래프로 표현하면 (그림 4)와 같다.

(그림 4) 각 그룹의 탐색 횟수

5. 결론

이 논문에서는 웹 미디어 시스템에서 이산 확률 해석과 사용자 프로파일 기반의 동적 적용 모델을 제안하였다.

확률적 해석 기법은 전체 응용 영역을 동적 적

용 콘텐츠 객체의 가중치 방향성 그래프로 생각한 논리적 탐색 그래프를 통해, 웹 미디어가 갖고 있는 정적 특성을 이용한 이산 확률 분포 함수이다.

이러한 확률적 해석은 웹 미디어 구조에서 사용자의 탐색 행위를 추적하여 얻을 수 있는 사용자 행위에 대한 잠재적 속성을 나타내는 데, 동적 적용 차원에서 사용자 행위에 관련된 속성을 사용자에게 관한 스테레오타입 모델에 연관하여 나타내었다.

사용자 행위는 사용자 행위 변수의 집합으로, 현재 프로파일, 현재 이산 확률 분포 함수, 최근 탐색한 패스, 최근 노드에서 시간 소모 등을 바탕으로 세 개의 이산 확률 분포 함수를 제안하였고, 이것을 바탕으로 사용자 분류를 하는 새로운 접근 방식을 제안하였다.

동적 적용 콘텐츠 객체는 사용자에게 제시될 페이지 개념의 단위로서, 동적 적용 프리젠테이션의 대상이 된다. 추상 개념과 응용 영역은 동적 적용 콘텐츠 객체를 방향성 그래프 구조로 나타내서 동적 탐색과 사용자 행위에 대한 평가를 할 수 있도록 하였다.

실험은 프로그래밍 영역에서 수행하였으며 초기에 구축된 스테레오타입의 사용자 모델을 실행 시에 사용자 탐색 행위를 추적하여 사용자의 행위에 따라 사용자에게 적합한 프로파일을 동적으로 적용시킴으로써, 학습 목표에 도달하는데 상당한 탐색 횟수를 줄일 수 있었다. 또한, 동적 적용 시스템이 사용자에게 효용성을 주고 있었으나 효율성을 저작 시 사용자의 스테레오타입 데이터와 추상 개념의 가중치 값의 정확성에 의존한다는 것을 확인하였다. 향후 연구로는, 정확한 가중치 값을 저작자가 쉽게 결정할 수 있도록 하는 시뮬레이터에 대한 연구가 필요하다.

참고 문헌

[1] Adaptive Hypertext and Hypermedia Home Page, <http://www.wis.win.tue.nl/ah/>.
 [2] Ardissono, L., and Goy, A.(2000), "Tailoring the Interaction With Users in Web Stores", in User Modeling and User-Adapted

Interaction, 10(4), Kluwer Academic Publishers.

[3] Baek, Y.T, Wang, C.J., and Lee, S.H.(2002), "Adaptive Hypermedia Educational System based on XML Technologies," ED-MEDIA 2002, AACE, USA.

[4] Brusilovsky, P.(2001), "Adaptive hypermedia", User Modeling and User Adapted Interaction 11 (1/2), pp. 87-110.

[5] Eklund, J. and Brusilovsky, P., & Schwarz, E.(1997), "Adaptive Textbooks on the WWW", <http://ausweb.scu.edu.au/proceedings/eklund/paper.html>.

[6] Emily Berk & Joseph Devlin(1998), Hypertext/Hypermedia Handbook, McGraw-Hill.

[7] Mario Cannataro and Andrea Pugliese(2001), "XAHM: an XML-based Adaptive Hypermedia Model and its Implementation," Hypertext'01, ACM.

[8] Mislevy, R. J. and Gitomer, D. H.(1996), "The Role of Probability-based Inference in an Intelligent Tutoring System," User Modeling and User-Adapted Interaction, 5(3-4).

[9] Murtagh, F., Tao, F.(2000), "Towards knowledge discovery from WWW log data", in Proceedings of the International Conference on Information Technology: Coding and Computing.

[10] Kobsa, A., Koenemann, J., Pohl, W.(2001), "Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships," The Knowledge Engineering Review 16(2), 111-155, Cambridge University Press.

백 영 태

1989 인하대학교
전자계산학과(이학사)

1993 인하대학교
전자계산공학과(공학석사)

2002 인하대학교 전자계산공학과

(공학박사)

1993~1998 대상정보기술(주) 정보통신연구소
선임연구원

2001. 3 멀티미디어기술사

1993~현재 김포대학 컴퓨터계열 조교수

관심분야: 하이퍼미디어시스템, 컴퓨터교육, 멀티
미디어컨텐츠개발

E-Mail: hanna@kimpo.ac.kr

이 세 훈

1985 인하대학교
전자계산학과(이학사)

1987 인하대학교
전자계산학과(이학석사)

1996 인하대학교 전자계산공학과

(공학박사)

1987~1990 해병대 분석 장교

1990~1993 (주)비트컴퓨터 기술연구소
선임연구원

1999.5 멀티미디어기술사

2001~2002 미국 뉴저지 공과대학(NJIT)
교환교수

1993~현재 인하공업전문대학 컴퓨터정보공학부
교수

관심분야: e-Learning, 하이퍼미디어시스템, 소프
트웨어공학, 분산객체컴퓨팅, XML/JAVA

E-Mail: seihoon@inhac.ac.kr