

적응형 웹미디어 교육 시스템을 위한 확률 모델

이재호^o 이윤수 윤경섭 왕창중

용인송담대학 안산공과대학 인하공업전문대학 인하대학교

jaeho1@ysec.ac.kr ystlee@ansantc.ac.kr ksyoon@inhac.ac.kr cjwang@inha.ac.kr

Probabilistic Model for Adaptive WebMedia Educational Systems

Jaeho Lee^o, Yoonsu Lee Kyeongseob Yoon Changjong Wang

Yongin Songdam College Ansan college Inha Technical College Inha Univ.

요약

이 논문에서는 웹 기반의 하이퍼미디어 교육시스템에서 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 모델링 한다. 제안한 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.

1. 서론

웹 미디어 시스템에서 사용자의 요구와 목적에 따른 프리젠테이션과 콘텐츠의 동적 적응 기능은 개인화는 중요한 요구 사항이 되고 있다[3,4]. 이러한 시스템은 전형적으로 대규모 하이퍼공간을 사용자가 자유롭게 탐색할 수 있도록 하고 있다. 그러나 이러한 웹 미디어의 풍부한 링크 구조는 일부 심각한 유용성 문제를 발생시키고 있다[1,5].

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 사용자에 대한 이해력 문제를 발생시키며, 개발자는 모든 페이지에 대해 사용자가 해당 페이지를 액세스할 때 사용자에 대해 가정을 예측해 시작해야한다. 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, 사용자에 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]. 개인화 시스템은 사용자의 동적 정보의 이용뿐만 아니라 사용자 탐색 행위를 관찰해 자동적으로 사용자 프로파일을 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 시스템과 확률과 베이지안 기법 등의 통계적 기법을 사용한 평가에 대한 연구들이 있다[6]. 이 논문에서는 웹 미디어 시스템에서 동적으로 개인화된 서비스를 할 수 있는 이산 확률 기법에 기반한 동적 적응 기법을 제안한다.

2. 확률적 해석기반의 동적 적응 모델

이 장에서는 웹 미디어 스키마 구조에서 아크의 가중치에 대

한 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 스테레오타입으로 구축하고, 실행 시에 사용자 행위를 모니터링해서 사용자의 행위에 따라 사용자 모델의 프로파일을 동적으로 조정하는 것이다.

이 연구에서는 전체 응용 영역을 동적 적응 콘텐츠 객체의 가중치 방향성 그래프로 생각한다. 이를 수식으로 나타내면, M 개의 다른 프로파일을 갖는 응용 영역과 동적 적응 콘텐츠 객체의 집합을 N이라 하면, 프로파일 k=1, ..., M에 대해 i ∈ N인 i 라는 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.

$$G_k = (N_k, E_k)$$

$$E = L_{1k} \cup L_{2k} \cup \dots \cup L_{Mk} = \bigcup_{k=1, \dots, M} L_{ik}$$

수식을 단순화하기 위해 가중치 방향성 그래프 G에서 각 프로파일에 대응하는 노드와 아크를 가중치 그래프 G_k(k=1, ..., M)라 하고, 이 G_k를 논리적 탐색 그래프라 한다. E_k에서 (i,j) 아크의 가중치를 w_k(i,j)라 하고 이것을 조건부 확률 P(j|k,i)로 정의하자. 즉, 프로파일 k에 속한 사용자가 i 노드에 도착해서 j 노드에 링크될 조건부 확률로 다음과 같이 정의된다.

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

그래프 G_k에서 패스 S를 순서 있는 노드의 집합으로 다음과 같이 정의한다.

프로파일 k에 속한 사용자 s가 패스 S를 따른 확률은 다음과 같다.

$$P_k^s = W_k(S_0, S_1) \times \dots \times W_k(S_{l-1}, S_l) \\ = \prod_{j=0}^{l-1} W_k(S_j, S_{j+1}) \\ = P_k^s(S_1|K, S_0) \times P_k^s(S_2|K, S_1) \times \dots \times P_k^s(S_l|K, S_{l-1})$$

여기서 S_j는 프로파일 k에 속하는 아크를 통하는 노드 i와 j 사이의 일반적인 패스이다. P_k는 패스 S에 속하는 아크에 확률

을 계산하는 것으로, 주어진 프로파일 k에 대해 i에서 j의 노드 사이의 패스 중 가장 짧은 패스를 S_{ij} 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.

$P_{ij}^k = \max_{(i,j) \in E_k} (P_{ij}^k)$ 이 논문에서는 웹 미디어 구조의 본질적 속성과 3개의 이산 확률 분포 함수(PDF; discrete Probability Distribution Function)로 모델을 제안한다.

$$\mu(k) = \frac{\sum_{(i,j) \in E_k} [\sum_{(i,j) \in E_k} P_{ij}^k \delta(k-i)]}{\sum_{(i,j) \in E_k} [N_{ij} \delta(k-i)]} \quad \nu(k) = \frac{\sum_{(i,j) \in E_k} [S_{ij} \delta(k-i)]}{\sum_{(i,j) \in E_k} S_{ij}} \quad \eta(k) = \frac{\sum_{(i,j) \in E_k} [N_{ij} \delta(k-i)]}{\sum_{(i,j) \in E_k} N_{ij}} \quad (식 1)$$

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.

$$s(k) = \frac{\beta_0 \mu(k) + \beta_1 \nu(k) + \beta_2 \eta(k)}{\beta_0 + \beta_1 + \beta_2} \quad (식 2)$$

여기서, $\mu(k)$ 와 $\nu(k)$ 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 k에 대한 높은

관련성을 나타내며, 따라서 $\beta_i > 0$ 이다.

3. 사용자의 동적 분류

이 연구에서 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 프로파일 이 사용자에게 맞는지 측정하는 것이다. 즉, 사용자 행위에 기반해서 시스템은 동적으로 사용자에 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-1을 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. 사용자 행위는 사용자 행위 변수의 집합으로 저장된다.

- 현재의 프로파일 K_0 .
- 현재 이산 확률 분포 함수 $A(k)(k=1, \dots, M)$ 은 각 프로파일 에 사용자의 소속 확률을 측정한다.
- 최근 탐색한 패스 $R = \{R_1, \dots, R_{i-1}, R_i\}$ 는 최근 방문한 노드를 포함한다. R_{i-1} 은 현재 노드이고 R_i 은 다음 노드이다. 최근 아크 (R_{i-1}, R_i, K_0) 는 사용자에 의해 선택된 출력 링크이다.
- 최근 노드에서 시간 소모는 $t(R_1), \dots, t(R_{i-1})$ 이다. 이것을 기초로 3개의 이산 확률 분포 함수를 정의한다.

$$c(k) = \frac{\sum_{(i,j) \in E_k} [P_{ij}^k \delta(k-i)]}{\sum_{(i,j) \in E_k} P_{ij}^k} \quad r(k) = \frac{\sum_{(i,j) \in E_k} [P_{R_{i-1}, R_i}^k \delta(k-i)]}{\sum_{(i,j) \in E_k} P_{R_{i-1}, R_i}^k}$$

$$d(k) = \frac{\sum_{(i,j) \in E_k} [D[i] \delta(k-i)]}{\sum_{(i,j) \in E_k} D[i]} \quad (식 3)$$

마지막으로 프로파일의 동적 관련성을 나타내는 가중치 중앙값은 다음 식으로 계산된다.

$$d'(k) = \frac{\alpha_0 c(k) + \alpha_1 r(k) + \alpha_2 d(k)}{\alpha_0 + \alpha_1 + \alpha_2} \quad (식 4)$$

성공을 기대하며, 이 프로파일은 이산하는 주요 아이디어는 사용자의 동적 행위를 결합하는데 있다. 사용자의 동적 행위는 주로 위상 기하학에 따른 웹 미디어 스키마 $s(k)$ 의 구조적 속성과 동합된 $d(k)$ 이다. <알고리즘 1>은 사용자 행동에 기반 한 새로운 확률 분포 함수 $A'(k)$ 를 계산한다.

<알고리즘 1> 새로운 확률 분포 함수 계산

<p>INPUT 이산 확률 분포 함수 $A(k), A_0(k), s(k)$ 최근 방문한 패스 $R = \{R_1, \dots, R_{i-1}, R_i\}$ 최근 방문한 노드에서 소비 시간, $t(R_1), \dots, t(R_{i-1})$</p> <p>OUTPUT 새로운 확률 분포 함수 $A'(k)$</p> <p>STEP 새로운 이산 확률 분포 함수 $d(k)$ 계산 새로운 이산 확률 분포 함수 $A'(k)$ 계산</p> <p>만약 $A'(k)$가 변화하면 $\Delta = \frac{\gamma_0 A_0(k) + \gamma_1 A(k) + \gamma_2 d(k) + \Delta \gamma_3 s(k)}{\gamma_0 + \gamma_1 + \gamma_2 + \Delta \gamma_3}$ 값이면 $\Delta = 0$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 k에 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

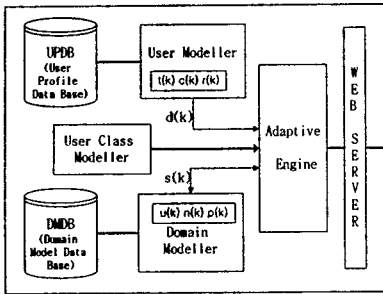
4. 실험 및 평가

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

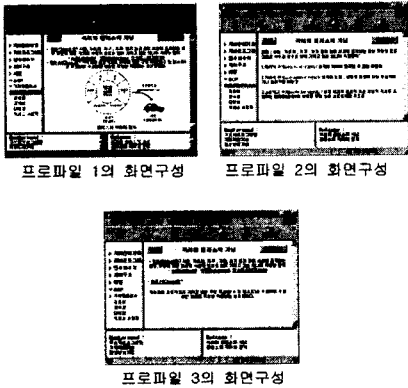
시스템은 (그림 1)과 같이 전체적으로 3개의 모델러와 2개의 DB로 구성된다. 영역 모델러는 웹 미디어의 정적 특성들을 모델링하며, 사용자 모델러는 동적 특성들을 모델링하게 된다. 동적 적응 엔진은 정적 특성과 동적 특성들을 이용하여, 사용자에게 가장 알맞은 최적의 프로파일을 동적으로 할당하게 된다. 시스템은 윈도우 2000 서버 환경에서 MS-SQL 2000, 웹 서버는 아파치 2.0, 웹 서버 스크립트 컨테이너는 JDK 1.4와 Tomcat 4.1, 문서의 동적 변환을 위한 COCOON 2.0을 이용하였다.

(그림 2)는 객체와 클래스의 개념에 대한 추상 개념 단계에서 각 프로파일별 링크에 가중치가 부가된 방향성 그래프로 구성된 형태이다. 각 프로파일은 자바 프로그래밍에서 초보자, 중급자, 전문가로 분류하고 학습 목적별로 분류하였다. 하나의 노드는 동적 적응 콘텐츠 객체를 의미하고 하나의 동적 적응 콘텐츠 객체 즉, 하나의 페이지에는 여러 개의 링크가 프로파

일련으로 다르게 구성되어 있다.



(그림 1) 동적 적응 시스템



(그림 2) 한 노드에서 구성된 프로파일별 화면

모든 프로파일 중 프로파일 1에서 보면 시작 노드 N1에서 목적 노드 N7로 갈 수 있는 일반적인 패스와 패스 S의 확률값은 (식 1)에 의하여 계산된 결과인 각 프로파일에 대한 하이퍼미디어의 구조적 속성값을 구한다. 현재 노드 N3를 방문한 사용자가 다음 링크로 노드 N4가 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 프로파일 사용자 확률로 저작자가 초기에 할당한 값으로 $A(1) = 0.33, A(2) = 0.33, A(3) = 0.33$ 로 놓자.

사용자가 속한 프로파일을 계산하는 주요 아이디어는 사용자의 동적 행위를 결합하는데 있다. <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 적응시킨다. 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.

5. 결론

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

확률적 해석 기법은 전체 응용 영역을 동적 적응 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 적응 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 적응 콘텐츠 객체는 사용자에게 제시될 페이지 개념의 단위로서, 동적 적응 프리젠테이션의 대상이 된다. 추상 개념과 응용 영역은 동적 적응 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 적응 시스템이 사용자에게 효율성을 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://wwwis.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] Brusilovsky, P.(2001), "Adaptive hypermedia", User Modeling and User Adapted Interaction 11 (1/2), pp. 87-110.

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

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

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