

강화 학습을 이용한 인터넷 문서 여과

서영우⁰, 장병탁
서울대학교 컴퓨터공학과

Internet Document Filtering Method Using a Reinforcement Learning

Dept. of Computer Engineering, Seoul National University

요 약

본 논문에서는 일반 사용자가 웹 문서공간에서 자신의 정보 요구에 적합한 정보를 찾을 때 맞게되는 부적절한 문서의 홍수에서 보다 적합한 문서를 여과시키는 방법을 제시한다. 제시된 방법은 적합성 최적-우선 탐색 방법으로 문서들을 검색하여 강화학습을 통해 학습된 사용자의 기호를 기준으로 문서들을 여과한다. 검색된 HTML문서는 사용자의 정보요구와의 적합성을 평가하는 하나 이상의 필드로 표현된다. 검색기간동안 학습에 이진트는 적합성검사에 사용되는 필드의 참여 비율을 조절하여 웹 문서 검색 에이전트의 검색정책을 수정한다. 제시된 결과에 대한 사용자의 적합성 평가는 사용자의 프로파일을 학습하고 웹 문서 검색 에이전트의 검색 정책에 대한 보상 또는 질책으로 사용된다. 실험은 기존의 웹 검색 엔진들과 제시된 방법을 동일한 topic에 대하여 각각의 검색결과에 대한 사용자의 만족도를 측정하였고 검색이 계속되는 동안 HTML문서에 대한 사용자의 기호를 학습하는 과정을 관찰하였다. 실험을 통해서 제시된 방법이 실제 환경에서 보다 나은 성능을 보였다.

1. 개 요

웹에서 이용 가능한 정보량이 증가함에 따라 사용자의 정보 요구에 보다 적합한 정보를 찾는 요구가 비례적으로 늘어가고 있다. 현재, 웹 문서정보에 접근하는 방법은 Alta vista, Excite등 같은 웹 검색 엔진을 이용하는 방법과 자신이 알고 있는 URL로부터 하이퍼 링크를 계속적으로 따라 가는 방법으로 크게 분류할 수 있다. 그러나, 첫 번째의 방법은 사용자의 부적절한 검색이 사용 또는 검색엔진의 부적절한 색인 방법으로 인하여 검색결과에 대부분이 사용자의 정보 요구에 대하여 연관성이 없는 내용인 단점이 있고, 두 번째의 경우는 적합한 정보를 찾는데 걸리는 시간과 노력이 너무 많이 든다는 단점이 있다.

본 논문에서는 강화학습을 이용하여 사용자 기호학습을 통한 웹 문서 검색과 검색된 문서들을 여과하는 방법을 제시한다. WAIR (Web Agents for Information Retrieval)¹⁾은 본 논문의 내용을 구현한 시스템으로 멀티-에이전트로 구성되어 있다. WAIR는 하나의 HTML문서를 4개의 HTML 필드-즉, Title, Header, Body, 그리고 Anchor text-로 간주한다. 각 필드는 검색된 문서와 사용자의 정보요구간의 적합성을 평가할 때 적합성-평가 요소로 사용된다. 이와 같이 WAIR는 HTML문서 구조를 반영하여 적합성평가를 한다. 검색이 계속되는 동안 제시된 검색결과에 대한 사용자의 적합성 평가는 사용자 자신의 프로파일을 학습하는데 사용되며, 사용자의 프로파일

이 생성되면 이 프로파일을 기준으로 검색된 문서가 여과된다.

2. 시스템구조와 웹 문서 여과 과정

WAIR의 동작 구조의 개요는 그림 1과 같다. WAIR는 사용자-인터페이스 에이전트, 웹 문서검색 에이전트,

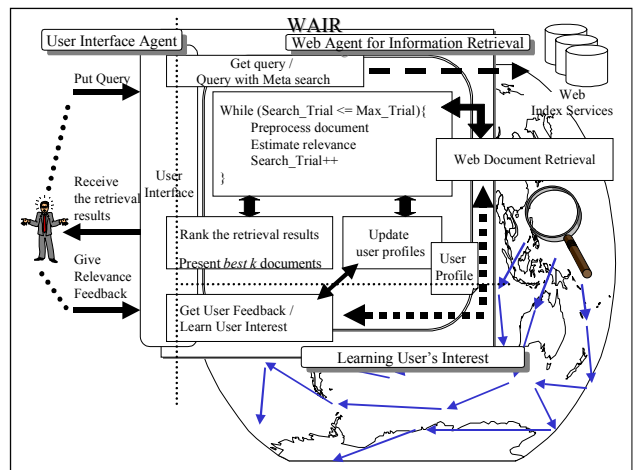


그림 1. WAIR 시스템 구조

감사의 글 : 본 연구는 대학기초 연구기술 지원사업(98-199)에 의해 지원되었음.

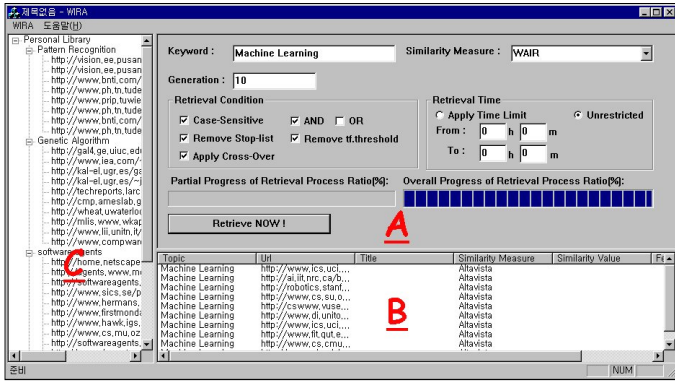


그림 2. WAIR의 사용자 인터페이스

그리고 기호 학습 에이전트로 구성된다. 사용자 인터페이스 에이전트는 사용자와 직접 상호 작용하는 에이전트로 그림 2와 같다. A부분은 사용자가 검색 조건과 질의어를 입력하고 검색 상태를 보는 부분이고 B부분은 검색 결과를 제시하는 부분이다. C부분은 사용자가 자신의 정보 요구에 대해 매우 적합하다는 평가를 내린 정보를 관리하는 개인용 도서관이다.

WAIR의 웹 문서 여과 과정은 그림 3의 의사코드로 요약할 수 있다. 사용자 인터페이스 에이전트로부터 질의어를 입력받아 메타-검색으로 검색된 HTML문서들은 단계 2-1부터 2-4까지의 과정을 거쳐 질의어와의 연관성이 평가된다. 이 과정은 초기에 주어진 종료조건을

Get a query Q from the user (Q : an user query).

Generate the initial set of candidate URLs, U^0 ,
using an existing search engines.

$U \rightarrow \{$.

Repeat following steps to build the document set U :

1. $U \leftarrow U \cup U^i$.
2. Repeat to generate U^i , i -th candidate set:
 - 2-1. Retrieve a document d in to U^i .
 - 2-2. Preprocess the document.
 - 2-3. Estimate the weight of each term in d .
 - 2-4. Estimate the relevance value of d to Q .
3. Sort the candidate set U^i in descending order of relevance value.
4. Get U^{i+1} by following k hyperlinks of the documents of high relevance value.

Rank U in descending order of relevance values.

Filter the best k HTML documents.

Learn user profile and retrieval strategy from the user relevance feedback.

그림 3. WAIR의 의사 코드(Pseudo code)

만족할 때까지 계속되는데 검색이 종료되면 적합성이 높은 순으로 k 개의 문서를 여과하여 사용자에게 제시한다. 각 HTML문서는 세부 단계 2-2에서 HTML태그와 불용어집합이 제거된 후 하나 이상의 단어와 그 단어의 속성들로 표현된다. 단어가 갖는 속성은 HTML문서 구조를 반영한 위치정보와 그 단어가 해당 문서에서 갖는 지역 중요도이다. 지역 중요도는 2-3에서 계산되는 가중치로 아래와 같다.

$$w_{d,t} = \frac{tf_{d,t}}{n_d}$$

여기에서 $w_{d,t}$ 는 d 번째 HTML문서의 t 번째 단어를, $tf_{d,t}$ 는 그 단어의 출현 빈도수를, n_d 는 d 번째 문서의 단어의 총수를 각각 나타낸다. 2-4에서는 아래의 식으로 각 문서와 사용자의 질의어와의 적합성을 평가한다.

$$F_{d,1} = \sum_{q_j \in V_T} w^{(T)}_j, F_{d,2} = \sum_{q_j \in V_H} w^{(H)}_j,$$

$$F_{d,3} = \sum_{q_j \in V_B} w^{(B)}_j, F_{d,4} = \sum_{q_j \in V_A} w^{(A)}_j$$

$$RV(d) = \sum_{i=1}^4 f_{t,i} \cdot F_{d,i} + \sum_{q_j \in TOP} w^{(TOP)}_j$$

3. 사용자 기호학습을 통한 웹 문서 여과

본 논문에서는 이미 기술한 바와 같이 하나의 HTML문서를 4개의 필드- Title $F^{(T)}$, header $F^{(H)}$, body $F^{(B)}$, 그리고 Anchor text $F^{(A)}$ -로 표현하며, 각 필드는 하나 이상의 단어와 그 단어의 가중치로 구성된다. 사용자의 질의어 $Q_i = (q_{i,1}, q_{i,2}, q_{i,3}, \dots, q_{i,n})$ HTML문서의 $w_{d,t}$ 와 비교 연산되어 검색되고 사용자 프로파일의 주제어 $tterm_{i,j}$ 와 비교 연산하여 여과된다. $tterm_{i,j}$ 은 사용자의 프로파일을 구성하는 i 번째 주제의 j 번째 단어이다. 이 주제어들은 n -CN(Contexts Neighbor)라는 특정 문장 내의 단어들의 상호 연관성을 고려하는 경험적인 방법이다. 즉, 사용자로부터 적합성을 높게 받은 HTML문서에서 Q_i 의 왼쪽 n 개와 오른쪽 n 개의 단어들을 선택하게 된다.

웹 문서검색 에이전트는 초기 메타-검색된 문서들에서 추출한 URL들을 출발점으로 하여 웹 문서공간을 탐색하는데, 검색방향은 적응형 최적 적합성-우선 검색을 사용한다. 적응형 최적 적합성-우선 검색이란 특정 사용자가 높은 적합성 평가를 내린 문서의 HTML구조와 유사한 구조를 갖는 HTML문서가 있는 쪽으로 웹 문서 검색의 방향을 조절하는 방법을 말한다. 즉, 웹 문서 검색 에이전트는 웹 문서 검색 정책에 의해 검색된 문서들을 사용자에게 제시하여 적합성평가라는 보상과 질책을 사용자로부터 받아 정책을 수정하면서 사용자의 동적인 기호를 학습한다.

이와 같이 학습자가 주어진 환경과 상호작용을 하며 실시간으로 학습하는 방법이 강화학습이다. 강화학습의 목적은 t 시간에 어떤 결정을 할 때 환경으로부터 오는 예상되는 누적 보상 값을 최대화시키는 최적의 정책을 찾는 것이다[1]. 사용자 기호학습 에이전트는 사용자의 적합성평가로부터 웹 문서 탐색정책을 학습하게 되며, 다음의 식으로 정의할 수 있다.

$$Q^\pi(s_t, a_t) = \sum_s P^a_{ss'} [R^a_{ss'} + \gamma \arg \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})]$$

$Q^\pi(s_t, a_t)$ 는 t 시간에 학습자의 행동을 결정하는 정책 π 로 어떤 행동을 취했을 때 그에 대한 가치를 계산하는 함수이다. 즉, 상태 s_t 에서 행동 a_t 를 취했을 때 정책 π 가 갖는 값에 대한 함수를 말한다. 본 논문과 같이 한정된(definite) MDP (Markov Decision Process)에서는 어떤 상태에서 어떤 행동을 취하는 것과 그 행동에 대한 환경의 반응값은 다음과 같이 어떤 확률값을 가지는 함수형태로 표현될 수 있다.

$$P^a_{ss'} = \Pr\{s_{t+1} = s' | s_t = s, a_t = a\}$$

$$R_{ss}^a = E\{r_{t+1} | s_t = s, a_t = a, s_{t+1} = s'\}$$

γ 는 감소요소(discount factors)로 t 시간 이후 정책이 현재 정책에 반영되는 정도를 표시한다. 본 논문에서 시간의 변화에 의존적인 상태 s_t 는 검색된 문서의 적합성 평가를 평가할 때 참여하는 각 필드의 비율로 다음과 같이 정의된다.

$$s_t = (f_{t,1}, f_{t,2}, f_{t,3}, f_{t,4}) \quad \text{with} \quad \sum_{i=1}^4 f_{t,i} = 1$$

이 때, 이 각 필드의 적합성 평가 참여 비율을 조절하는 것을 행동으로 정의하면 다음과 같다.

$$a_t = (\Delta f_{t,1}, \Delta f_{t,2}, \Delta f_{t,3}, \Delta f_{t,4}) \quad \text{with} \quad \sum_{i=1}^4 \Delta f_{t,i} = 1$$

일반적으로 상태 s 를 s' 으로 변경시키는 행동은 학습자의 행동의 탐험(exploration)과 개척(exploitation)을 조절하기 위해서 ϵ -greedy 방법을 사용하지만 본 논문에서는 사용자의 적합성 평가를 기반으로 아래와 같이 greedy 방법을 사용했다.

$$\Delta f_{t,i} = \begin{cases} f_{t,i} + \Delta f_{t-1,i} & \text{if } r_{t-1} \in [0.7, 1.0] \\ f_{t,i} - \Delta f_{t-1,i} & \text{if } r_{t-1} \in [0.0, 0.3] \\ f_{t-1,i} & \text{otherwise} \end{cases}$$

$$\Delta f_{t-1,i} = f_{t-1,i} - f_{(t-1)-\Delta t,i}$$

4. 실험 및 결과

실험은 목적은 WAIR시스템이 사용자에게 제시된 결과에 대한 사용자의 만족도가 점진적으로 증가하는 것[표 2]과 본 논문에서 제시한 방식이 기존의 웹 검색엔진에 비해 사용자 관점에서 어느 정도의 만족도 차이를 보이는지 관찰하는데 있다[그림 3]. 사용자의 적합성 평가는 0에서 1까지 11계급으로 각 계급의 차는 0.1이며, 감소 요소 γ 의 값은 0.9로 $t+1$ 시간의 정책만을 반영하였다. 10명의 실험자가 아래 표 1의 질의문을 이용하여 실험에 참석하였다.

질의문 번호	질의문
1	Machine & Learning & Introduction
2	Reinforcement & Learning & Agent
3	British & Museum travel tour
4	Machine & Learning DW DM KDD
5	Collaborative & filtering & EC
6	Software & agent & workshop
7	Genetic Algorithm & relevance & feedback
8	Soccer & World & Cup & Ronaldo
9	Leonardo da Vinci uffizi Museum
10	Beatles & History Picture

표 1. 실험 자료, &:And search, |:Or search

초기 적합성 평가 요소값은 0.25, 변화량은 0.25로 크기로 각각 정의하여 실험을 실시하였으며, 사용자가 명시적으로 적합성 평가를 내리지 않았을 때는 최종 결과로 제시된 검색문서가 가지는 적합성 값의 상대적인 수치를 적합성 평가값으로 계산하였다.

5. 결론

본 논문에서는 개인 사용자가 자신의 정보 요구에 보다 적합한 웹 문서를 찾기 위한 웹 문서 여과 방법을 제시하였다. 실험을 통해서 제시된 방법이 사용자의 특정 주제에 대한 관심에 보다 적절한 문서들을 제시하는 것을 보았으며 검색시도가 계속될수록 사용자의 기호에 적응해 가는 것을 볼 수 있었다. 향후과제로는 웹 문서에 대한 사용자

의 기호를 학습하는데 HTML문서의 구조를 좀더 많이 반영하는 것과 문서의 적합성을 평가할 때 웹 문서들간의 연결구조를 고려하는 것이 있다.

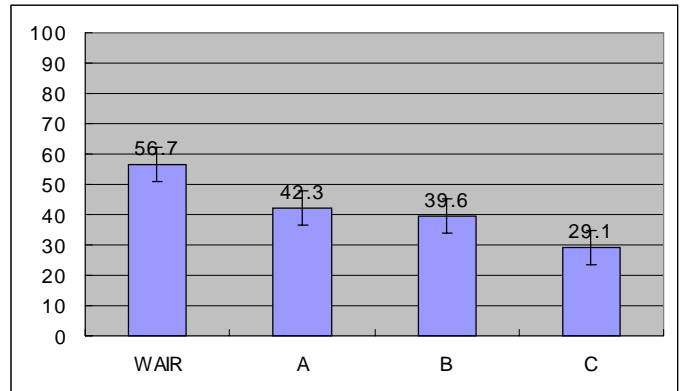


그림 3. 제시된 문서에 대한 사용자 만족도 비교, x축: A,B,C 기준 검색엔진, y축: 평균 적합성평가

Feedback Trial Query no.	average/standard deviation/ positive/neural/negative			
	1	5	10	15
1	55/28.8 4/3/3	68/16.9 5/4/1	76/24.1 8/1/1	73/27.5 9/0/1
2	58/26.6 4/4/2	55/27.6 4/3/3	84/20.7 9/1/0	67/25.1 7/2/1
3	69/36.3 6/2/2	54/35 6/1/3	70/36.2 8/0/2	67/27.1 6/3/1
4	58/31.2 4/3/3	43/25.8 2/6/2	43/47.6 5/0/5	64/30.6 6/2/2
5	67/39.7 6/0/4	61/18.35 4/6/0	43/47.2 4/0/6	67/18.2 7/3/0
6	48/30.5 2/6/2	65/15.8 7/2/1	71/27.2 8/1/1	74/27.6 7/2/1
7	41/33.1 2/2/6	70/18.3 7/3/0	52/42.4 5/1/4	69/28.5 7/2/1
8	53/39.7 4/3/3	70/27.5 7/2/1	73/39.2 7/1/2	79/31.1 7/2/1
9	45/35.7 4/2/4	66/21.7 5/4/1	56/44.8 5/1/4	73/29.8 8/1/1
10	763/24.7 6/3/1	45/37.2 4/0/6	39/47 4/1/6	72/28.6 7/2/1

표 2. 사용자 적합성 평가의 분포

참고문헌

- [1] Tom M. Mitchell, *Machine Learning*, McGraw-Hill, pp.369-390, 1997.
- [2] A. Falk and Ing-Marie Jansson, PAWS: An agent for WWW-Retrieval and filtering, In *Practical Application of Intelligent Agents and Multi-agents Technology '96*, pp169-179, 1996.
- [3] F. Menczer, ARACHIND: Adaptive retrieval agent choosing heuristic neighborhoods for information discovery, In *International Conference on Machine Learning '97*, pp227-235, 1997.
- [4] G. Salton, *Automatic Text Processing*, Addison-Wesley, 1989.