강화학습을 이용한 부정적 연관성 피드백

손기준○ 이재안 이상조
경북대학교 컴퓨터공학과
kijunsonO@msn.com, jalee@sejong.knu.ac.kr sjlee@bh.knu.ac.kr

Negative Relative Feedback Using Reinforcement Learning

Kijun Son○ Jaean Lee Sangjo Lee
Dept. of Computer Engineering, Kyungpook National University

요 약

문서 여과 시스템은 사용자의 정보요구를 기준으로 문서들을 선별하여 제시한다. 사용자의 정보요구는 하나 이상의 단어들로 구성된 프로파일로 표현이 되며, 문서의 여과 과정 동안에 발생하는 사용자의 연관성 평가를 통해 구체적인 내용으로 변할 수 있다. 기존 연구의 경우 사용자는 자신이 직접 연관성 평가에 참여하여 평가 정보를 입력하고, 사용자가 평가한 긍정적 피드백 정보를 이용하여 사용자 프로파일을 학습한다. 본 연구는 사용자가 평가한 긍정적 연관성 피드백 뿐만 아니라 부정적 연관성 피드백을 함께 이용한 사용자 프로파일 학습 방법을 제안한다. 제안된 방법과, 대표적인 연관성 피드백 방법인 Rocchio 방법과의 성능을 측정하기 위해 네 가지 토픽에 대하여 여과를 수행하였다. 실험한 결과 부정적 연관성 피드백 정보를 이용하였을 경우 Rocchio 방법 보다는 6% 더 성능이 높은 것을 볼 수 있었다. 실험결과 부정적 평가를 받은 문서를 이용하여 사용자가 선호하지 않는 문서를 제거함으로써 여과 시스템의 성능을 향상 시킬 수 있었다.

1. 서 론

인터넷의 발달로 전자 정보의 양이 증가함에 따라 사용자의 정보요구에 보다 적합한 정보를 찾는 요구가 비례적으로 늘어가고 있다. 사용자는 정보의 양이 증가할수록 모든 정보를 확인해 가며 자신이 원하는 정보를 찾아보기가 힘들어지게 된다. 따라서 컴퓨터가 사용자를 대신하여 양질의 정보를 여과해주는 여과 시스템은 매우 유용한 도구가 될 것이다.

여과 시스템은 이러한 방대한 정보 집합 속에서 사용자의 정보요구와 선호도에 대한 관련정보를 제시함으로써 사용자를 돕도록 의도된 지능적인 시스템이라 정의할 수 있다[1]. 개인화된 지능적 정보 에이전트는 사용자의 정보요구와 선호도를 직접적 또는 간접적으로 학습하여 사용자 프로파일을 구축하게 된다[3].

사용자의 프로파일을 구축하기 위하여 사용자의 평가를 이용한 연관성 피드백 방법을 이용한다[5,6,7]. 즉 연관성 피드백 방법은 시스템으로부터 사용자가 일차 여과된 문서 중에서 적합문서와 부적합 문서를 판단한 후 사용자가 관련 있다고 생각되는 문서의 키워드를 사용자의 프로파일에 추가하여 사용자의 선호도를 변경한다[5,6,7]. 즉 사용자의 평가만이 수치적으로 주어지는 환경이다. 이러한 방법은 사용자가 선택한 적합문서의 키워드를 이용하여 사용자 프로파일을 구축하고, 이를 이용하여 사용자가 필요로 하는 문서를 여과한다. 하지만 여과되는 문서의 양이 많아지고, 모호한 단어가 출현 하였을 경우 부적합 문서가 섞일 가능성이 크다. 이에 따라 본 연구에서는 강화학습을 연관성 피드백 관점에서 분석하고 부정적인 연관성 피드백을 이용하여 사용자가 선호하지 않는 문서를 여과 과정에서 제거한다. 또한 사용자의 선호도 학습을 위해 사용자가 평가한 부정적 관련성 평가 값을 이용하여 사용자 프로파일을 구성한다. 이렇게 구성된 사용자의 프로파일은 강화학습[3]을 이용하여 학습을 수행하며 학습 과정이 끝난 후 사용자의 정보요구에 해당하는 문서를 여과하게 된다.

본 논문의 구성은 다음과 같다. 2장은 관련연구로서 연관성 피드백 방법과 강화학습에 대하여 살펴보고, 3장에서는 본 연구에서 제안한 부정적 연관성 피드백을 이용한 사용자 선호도 학습에 대하여 설명한다. 4장에서는 실험을 통한 결과를 설명하고, 5장에서는 결론 및 향후 연구 과제를 기술한다.

2. 관련 연구

2.1 연관성 피드백

연관성 피드백은 검색 분야에서 오랜 기간 연구가 되어 왔으며 피드백이 없을 경우에 비해 많은 성능의 향상을 가져오는 것으로 알려져 왔다[6].

문서여과에서 강화학습을 이용한 사용자 프로파일 학

습에 관한 연구로는 [1][9][10]이 있다. [1]의 연구에서는 사용자로부터 주어지는 명시적인 평가와 사용자의 행위 분석을 통한 묵시적인 정보를 이용하여 사용자의 선호도를 학습하고 문서를 여과한다. 하지만 별개의 주제에 대한 복수의 선호도를 표현 할 수는 없다. [9]의 연구에서는 사용자의 문서에 대한 관심도에서 사용자가 선호하는 용어를 추출 이를 강화학습으로 학습하는 방법을 사용한다. [10]의 연구에서는 사용자의 모델링을 위해 강화학습 기반의 새로운 연관성 피드백 알고리즘을 제안하였다. 위의 연구들에서는 긍정적인 정보만을 이용하여 프로파일을 학습함으로써 여과된 문서에 노이즈가 들어갈수 있다. 이에 따라 본 연구에서는 사용자가 평가한 부정적인 관련성 값도 프로파일의 학습에 이용한다.

연관성 피드백 모델에서 각 문서들은 사용자에 의해 적합도 평가값이 주어지게 되고, 문서와 그에 해당하는 적합도 평가 값으로부터 보다 평가값이 높은 문서들을 얻을 수 있도록 질의문이 수정된다. 가장 대표적인 연관성 피드백 방법은 Rocchio[6]의 알고리즘으로, 질의문은 식 (1)과 같이 수정된다. 이 방법은 전체 문서를 한꺼번에 처리하는 배치 알고리즘으로써 하나의 상태에서 한번의 행동을 통해 한번 학습하게 되기 때문에 여러 번의 시행을 통해 학습을 하지는 못한다. 즉 긍정적 예제의 속성은 사용자 프로파일에 추가되고 부정적 예제의 속성은 사용자 프로파일로부터 제거되는 과정을 거쳐 학습을 수행한다.

$$Q_1 = Q_0 + \beta\sum_{i=1}^{n_1}\frac{R_i}{n_1} - \gamma\sum_{i=1}^{n_2}\frac{S_i}{n_2} \quad (1)$$

식 (1)에서 Q_0는 초기 질의 벡터이고 Q_1은 확장된 질의 벡터를 나타내며, R_i는 관련문서 i의 벡터를 S_i는 비관련 문서 i의 벡터이다. 이때 n_1, n_2는 선택된 관련/비관련 문서의 수를 나타내며 β,γ는 관련/비관련 용어의 중요도 조정 가중치이다. [8]의 연구에서 β,γ를 0.75, 0.25의 값을 이용하였을 경우 가장 좋은 피드백 결과를 나타내었다. 하지만 본 연구에서는 관련/비관련 문서이 동일한 가중치인 β,γ는 0.5의 가중치를 사용한다.

2.2 강화학습

강화학습은 감독학습과 무감독학습의 중간적인 특성을 띠고 있으며 동적인 환경에서 시행착오를 거치면서 환경으로부터 주어진 보상을 최대화하기 위한 학습 방법이다 [3,4]. 환경(environment)은 주로 상태로 표현되며, 행위자(agent)는 적절한 정책(policy)에 따라 행동을 취하게 된다. 이때 환경은 행위자에게 행동에 대한 보상을 주게 된다. 강화학습 에이전트가 t시각에 행동 a_t를 취하면 행동에 대한 보상 r_t가 환경으로부터 주어진다. 그리고 행동에 의하여 상태 s_t가 s_{t+1}로 변화된다. 이러한 행위자와 환경과의 상호작용은 그림 1과 같다.

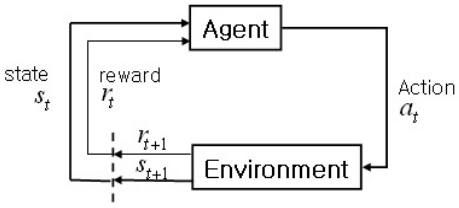


그림 1. 강화학습에서 에이전트와 환경과의 상호작용

2.3 Q-Learning

Q-Learning은 상태 s에 대한 행동 a의 가치함수 $Q(s,a)$를 기반으로 각 단계마다 에이전트는 최대 가치함수 값을 가진 행동 a를 선택한다. Q-Learning은 미래 보상값을 할인하기 때문에 장기간 동안 보상값들을 유지하는 행동보다는 단기간에 끝나는 행동을 선호한다. Q-Learning은 매 번의 행동 후 $Q(s,a)$ 값은 다음 규칙으로 갱신한다.

$$\widehat{Q}(s,a) \leftarrow r + \gamma \max_{a'} \widehat{Q}(s',a') \quad (2)$$

r은 행동 a의 수행에 대한 보상이며, γ는 Q가치함수 값의 정학도를 보장하는 할인상수 이다. 이것은 먼저 행한 행동에 대한 보상값은 높은 값으로 정하고, 같은 행동에 대하여 시간이 흐름에 따라 그 보상값을 감소 시키는 것이다. 즉 할인 상수가 γ라면 어떤 행동이 n번째 단계에 행해졌을 때의 보상값은 초기 보상값에 γ^n을 곱한 값이 된다.

3. 강화 학습과 부정적 연관성 평가를 이용한 사용자 선호도 학습

사용자 선호도 학습의 목적은 문서 여과 과정에서 사용자의 정보요구를 명확히 반영하는 데 있다. 즉 사용자에게 유익한 관련 정보만을 선별적으로 제공하여야 한다. 이러한 특징을 지니기 위해서는 무엇보다도 사용자의 선호도를 잘 표현할 수 있는 프로파일의 정확한 구축이 필수 조건이며, 이를 위해 사용자의 연관성 피드백을 이용한다.

정보검색의 대표적인 모델인 벡터 공간 모델에서 사용자의 피드백을 이용할 때 사용자가 명시적으로 제공한 연관성 정보만을 이용하여 사용자의 프로파일을 학습하는 방법을 사용한다[5]. 사용자가 관심을 가지는 문서 중 사용자가 평가한 적합문서만을 학습에 이용한다. 이

에 따라 여과 과정 중 여과되는 문서에 부적합문서가 포함 된다. 즉 사용자가 평가하지 않은 부적합문서는 학습에 이용되지 않으므로 여과 과정 중 부적합 문서가 포함될 수 있다. 이때 사용자가 관심을 가지지 않는 부적합 문서도 함께 학습에 이용함으로써 사용자의 관심 영역을 명확히 구분할 수 있다.

예를 들면, 사용자가 '애플(컴퓨터)'에 관심을 가지고 있을 경우 사용자가 평가한 적합 문서를 이용하여 학습 과정을 수행하고 이어지는 다음 문서 스트림으로부터 여과를 수행하게 된다. 이때 사용자가 평가한 적합 문서는 '애플컴퓨터'에 관한 문서들로 구성 되어 있었지만 부정적인 평가 정보를 이용하지 않음으로 인하여 여과 과정에 '애플(과일)'에 관련된 문서가 여과될 수 있다. 이에 따라 사용자가 평가하지 않은 부적합 문서를 학습에 이용함으로써 사용자에게 제출되어지는 부적합 문서의 양을 줄일 수 있다.

본 논문에서 학습 에이전트의 상태 s_t는 사용자 프로파일에 있는 용어들이며, 행동 a_t는 문서에 대한 사용자의 평가 정보에 의하여 가장 높은 관련성 평가값을 받은 문서를 선택(ε-greedy[3])하고. 이 문서에 나타난 용어를 이용하여 프로파일을 구성하는 것으로 (3)에 따라 갱신된다. 처음 생성되는 용어의 가중치는 tf값을 사용한다.

$$w_{pk} \leftarrow w_{pk} + r_i \tag{3}$$

식 (3)에서 r_i는 여과된 문서 i에 대한 사용자의 관련성 평가값이며, w_{pk}는 프로파일 p의 k번째 단어의 가중치이다.

학습 에이전트의 목표는 사용자가 관심을 보인 문서에 가중치를 부여하여 용어를 선별한 후 최적의 프로파일을 유지하는 것이다. 사용자로부터 주어지는 문서 D_i에 대한 관련성 피드백은 스칼라 값을 가지며 식 (4)에 의해서 구해진다.

$$R(i) = \sum f_v(i) \tag{4}$$

위의 식에서 f_v는 사용자의 긍정적, 부정적 연관성 피드백에 따라 주어지는 가중치이며, 여과된 문서에 대한 사용자의 선호도가 프로파일에 반영된다. 즉, 프로파일 p의 k번 단어가 적합성 평가를 받은 문서 D_i내에 있으면 가중치 w_{pk}에 사용자가 제공한 관련성 평가값을 α만큼의 반영한다. 이때 사용자가 제공한 관련성 평가는 5단계로 나누어 사용한다. 긍정적인 평가는 '보통', '관련 있음', '아주중요'의 3단계로 나누어 사용하며 1~3의 평가값을 사용하며, 부정적인 평가는 '관련 없음'과 '응답 없음'의 2단계로 나누고 0~-1의 평가값을 사용한다. 이 평가값을 이용하여 단어의 가중치를 변경하게 된다. 사용자의 프로파일은 식 (4)에서 주어지는 연관성 피드백 값 r_i을 이용하여 식 (5)과 같이 사용자 프로파일을 갱신한다.

$$w_{pk} = w_{pk} + \alpha r_i \tag{5}$$

사용자 선호도 학습 과정은 사용자로부터 연관성 피드백을 받고, 이 정보를 이용하여 사용자 프로파일을 학습한다. 즉, 사용자의 연관성 피드백을 이용하여 프로파일을 수정하며, 사용자의 선호도를 잘 반영하게 하는 것이 프로파일 학습의 목적이 된다. 사용자로부터 주어진 긍정적, 부정적 연관성 피드백을 이용한 프로파일의 학습 과정은 그림 2로 요약할 수 있다. 사용자의 프로파일은 하나 이상의 주제(Topic)와 해당 주제어를 포함하는 단어들의 가중치로 구성된다. 사용자 프로파일 w_p는 주제어 p에 대한 프로파일로 식 (6)과 같이 하나 이상의 단어들과 각 단어의 가중치 벡터로 표현된다.

$$w_p = (w_{p1}, w_{p2}, \ldots, w_{pk}, \ldots, w_{pn}) \tag{6}$$

식 (6)에서 w_{pk}는 k번째 단어의 가중치이며 사용자의 프로파일을 표현하기 위해 n개의 단어를 이용한다.

1. 사용자 프로파일 p을 초기화. $set\ t \leftarrow 0$
2. Q(s,a)를 초기화, 모든 s, a에 대하여
2. 사용자의 행동이 끝날대 까지 반복.
   2.1 $s \leftarrow$ 현재 상태.
   2.2 ε-greedy 정책에 의해 행동 선택.
   2.3 사용자로부터 관련성 평가를 받음.
   2.4 보상값을 식 (4)로 계산.
   2.5 용어 가중치 변경 (3).
3. 사용자 프로파일 갱신. 식 (5).
4. $t \leftarrow t+1$, 단계 2로 이동.

그림 2. 사용자 프로파일 학습 과정

4. 실험 및 분석

실험에 사용된 데이터는 중앙일보 신문기사를 대상으로 수집된 총 770개의 문서를 사용한다. 이 문서의 구성은 중앙일보 웹 사이트에서 제공되는 11개의 카테고리에서 각각 70개의 문서를 수집 사용하였고, 카테고리는 정치, 경제, 스포츠, 문화 등으로 구성되어 있다. 여과 시스템의 성능을 평가하기 위해 네 가지 토픽에 대하여 성능을 평가한다. 실험에 사용한 네 가지 토픽은 '올림픽', '유가(油價)', '연예계', '문화'이다. 실험 조건상 사용자

선호도 변화는 없는 것으로 가정하였다.

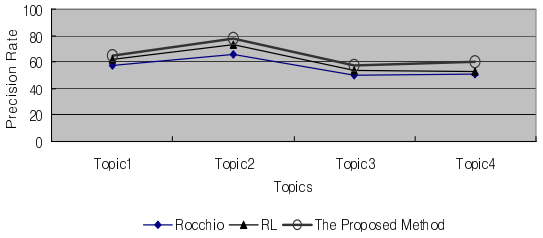
문서들은 전처리 과정을 통하여 TF벡터로 표현하고, 이 문서들에 나오는 단어들과 사용자의 프로파일에 나오는 단어들을 이용하여 보상값을 구한다. 사용자의 정보 요구에 맞는 정보를 여과하기 위해 사용자 프로파일을 바탕으로 보상을 책정한다. 즉 현재의 문서가 얼마나 사용자 프로파일에 유사한지의 값을 바로 보상으로 줄 수가 있다.

실험은 각 토픽별로 Rocchio, RL, 제안한 방법을 이용한 경우에 대하여 여과의 성능을 평가하였다. 여과 결과의 평가 척도로는 정보 검색에서 널리 사용되고 있는 정확률과 재현율을 이용하여 평가 하였으며 사용자 프로파일과 문서 간의 유사도를 측정하기 위해 코사인 유사도 측정법을 사용하였다. 평가 대상 문서는 Rocchio, RL 방법을 이용하여 여과된 문서 중 상위에 랭크된 10, 20, 30개의 문서를 대상으로 한다.

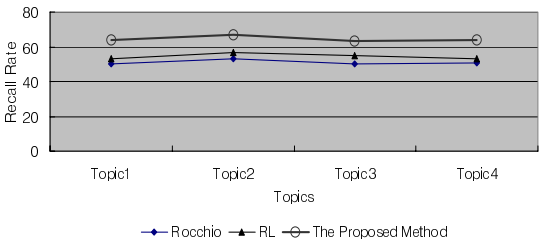
실험한 결과 상위 10개의 문서에 대한 여과 성능은 그림 3, 과 같으며, 상위 20개, 30개의 문서에 대한 여과 성능은 그림 4, 5와 같다. 그 결과 제안한 방법이 Rocchio 방법 보다는 6~10% 성능이 높은 것으로 나타났으며, RL과 긍정적 관련성 평가만을 이용한 경우 보다 제안한 방법의 경우 2~3% 성능 향상을 볼 수 있었다.

실험 결과에서 볼 수 있듯이 사용자로부터 받은 긍정적 관련성 평가와 부정적 관련성 평가를 받은 문서를 함께 이용한 경우 여과의 성능이 향상되었다. 또한, 여과되는 문서의 수가 많아지면 성능이 향상되는 것을 알 수 있다.

일반적인 주제를 다루는 Topic 3과 Topic 4는 큰 성능의 향상을 보이지 않았다. 하지만 Topic 2는 다른 토픽들 보다 더 나은 성능을 보이고 있다. 그 이유는 Topic 2에 등장한 어휘들이 다른 토픽에 나타난 어휘들보다 분별력이 높기 때문인 것으로 추정된다. 즉 그 주제에 나타난 어휘가 다른 주제에는 자주 사용되지 않기 때문이다. 예를 들면, 유가(油價)관련 문서에서 유가, 휘발유, 등유, OPEC, 두바이유와 같은 어휘들은 다른 주제에는 자주 사용되지 않는다. 따라서 이러한 어휘들이 유가관련 문서에 국한되어 나타남으로써 여과의 성능을 높여 준다.

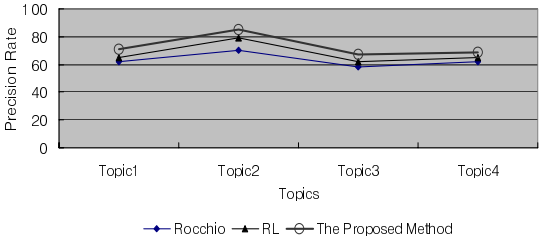


(a) 정확률

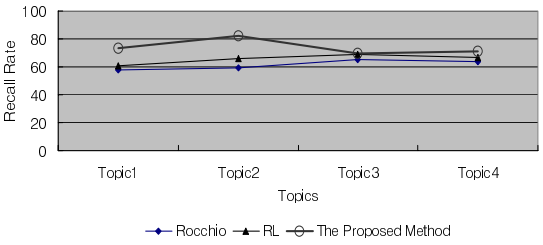


(b) 재현율

그림 3. 상위 10개의 문서에 대한 여과 성능

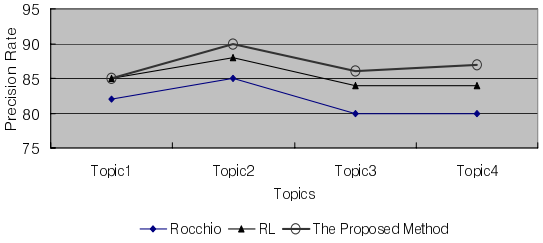


(a) 정확률

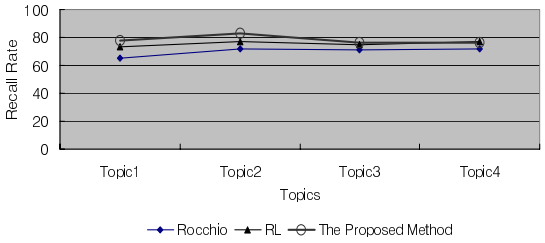


(b) 재현율

그림 4. 상위 20개의 문서에 대한 여과 성능



(a) 정확률



(b) 재현율

그림 5. 상위 30개의 문서에 대한 여과 성능

5. 결론

본 논문에서는 강화학습을 연관성 피드백 방법의 관점에서 분석하고 이를 이용하여 사용자 프로파일을 학습하는 방법을 제시하였다. 기존의 수동적인 문서 여과 시스템은 많은 양의 문서들 중 사용자가 선호하는 문서를 선택해 주지만, 많은 양의 문서에서 사용자에게 매번 선택을 요구하는 것은 무리가 있다. 이에 따라 강화학습을 이용하여 사용자의 프로파일을 학습한 후 사용자의 관심에 맞는 문서를 여과해주는 여과 시스템으로 사용자의 정보 요구에 응할 수 있는 장점을 가진다.

실험한 결과, 제시된 방법이 특정 주제에 대한 관심에 보다 적절한 문서들을 제시하는 것을 보았다. 즉 사용자가 제시한 부정적 연관성 피드백을 함께 이용하여 사용자 프로파일을 학습하였을 경우 긍정적 연관성 피드백을 이용한 경우 보다 나은 성능을 보였다. 사용자들이 많은 양의 기사를 모두 읽고 자신의 관심 주제를 오류 없이 모두 명시하는 것은 쉽지 않은 일이므로, 제안된 방법이 어느 정도의 오류 및 누락과 관계없이 동작할 수 있는 이러한 특징은, 신문기사 여과 서비스와 같은 실용적인 용도의 여과를 위해 바람직한 것으로 보인다.

향후 과제로는 사용자의 관심이동 정보를 반영한 프로파일의 학습 방법에 대한 연구가 필요하다. 이와 더불어 웹 환경에서 사용자의 프로파일의 공유를 위해 온톨로지를 활용한 사용자의 선호도 학습 방법에 대한 연구가 필요하다.

[참고문헌]

[1] Seo, Y., Zang, B., “Personalized Web Document Filtering Using Reinforcement Learning,” Applied Artificial Intelligence, Vol. 15(7), pp. 665-685, 2001.

[2] M. Balabanovic, Y. Shoham, "Learning Information Retrieval Agent: Experiments with Automated Web Browsing," In Proceeding of the AAAI Spring Symposium on Information Gathering, Stanford, CA, March 1995.

[3] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

[4] L. P. Kaelbling, M. L. L. Littman and A. W. Moore, "Reinforcement Learning: A Survey," Journal of AI Research, vol. 4, pp. 237-285, 1996.

[5] G. Salton, M. J. McGill, Introduction to modern information retrieval, McGraw Hill, 1983.

[6] Rocchio J., "Relevance feedback information retrieval," In Gerard Salton, ed., The Smart retrieval system experiments in automatic document proc. Prentice-Hall, Englewood, NJ, 1971.

[7] Huang, X., Huang, Y., Wen, M., An, A., Liu, Y and Poon, J. "Applying Data Mining to Pseudo-Relevance Feedback for High Performance Text Retrieval", The Proceedings of the 2006 IEEE International Conference on Data Mining(ICDM'06), Hong Kong, China, Sep 2006.

[8] Salton, G. and Buckley, C., "Improving retrieval performance by relevance feedback," Journal of American Society for Information Science, 41:288-297, 1990.

[9] 김영란, 한현구, “강화학습 기반 사용자 프로파일 학습”, 한국정보과학회 가을 학술발표논문집, Vol. 29. N0. 2, pp. 325-327, 2002

[10] 이승준, 장병탁, “강화학습을 사용한 연관성 피드백“, 한국정보과학회 가을 학술발표논문집, Vol. 29. N0. 2, pp. 280-282, 2002.