

관련성과 강화학습을 이용한 신문기사 추천시스템

상대중^o 손기준[†] 박미성[†] 이상조[†]
 경북대학교 산업대학원 컴퓨터공학전공^o, 경북대학교 컴퓨터공학과[†]
 tjsang@lgcns.com^o, kjson@sejong.knu.ac.kr[†], {mspark, sjlee}@knu.ac.kr[†]

News Article Recommender System By Relevance and Reinforcement Learning

T. J. Sang^o, K. J. Son[†], M. S. Park, S. J. Lee[†]
 {Computer Engineering Major Graduate School of Industry^o, Dept. of Computer Engineering[†]}
 Kyungpook National University

요약

추천 시스템은 양질의 정보를 추천하기 위해서 사용자의 관심도를 반영해야 한다. 이를 위해 본 연구에서는 강화학습과 관련 정보, 비관련 정보를 모두 이용하는 피드백 방법을 결합하였다. 사용자의 문서에 대한 평가를 평가값으로 사용하여 사용자가 선호하는 용어와 선호하지 않는 용어를 추출하고, 이를 이용해 사용자 프로파일을 강화학습으로 학습하게 된다. 제안된 방법으로 신문기사 추천시스템에 적용하여 실험한 결과, 관련 정보와 비관련 정보를 함께 사용한 방법이 기존의 관련 정보만을 사용한 방법보다 더 나은 성능을 보였다.

1. 서론

인터넷의 발달로 인하여 전자 정보의 양이 증가하고 있다. 사용자는 정보의 양이 증가할수록 모든 정보를 확인해 가며 보기가 힘들어지게 된다. 따라서 컴퓨터가 사용자를 대신하여 사용자가 원하는 양질의 정보를 추천해주는 추천 시스템은 매우 유용한 도구가 될 것이다. 대규모의 전자문서로부터 원하는 정보를 정확히 찾고 그 텍스트의 내용을 분석하여 사용자가 요구하는 정확한 정보를 제공할 수 있는 정보 기술의 개발에 유용하게 이용될 수 있다.

추천 시스템의 기술을 향상시키는 방법들이 여러 가지지만 그 중에서도 사용자의 관심도를 반영하여 추천 시스템의 효율을 높이고 개인화된 추천 서비스를 제시할 수 있는 방법이 먼저 고찰되어야 한다. 개인화된 지능적 정보 에이전트는 사용자의 정보 요구와 선호도를 직접 또는 간접적으로 학습하여 사용자 프로파일을 구축하게 된다 [1].

이때 사용자의 선호도를 학습하기 위해 사용자의 평가를 이용한 연관성 피드백 방법을 이용하기도 한다. 이러한 대부분의 시스템은 사용자가 제시한 관련성 정보만을 이용하여 프로파일을 학습한다. 이에 본 논문에서는 사용자가 제시한 관련성 정보 뿐만 아니라 비관련성 정보를 함께 이용한 사용자 프로파일 학습방법을 제시하고, 실험을 통하여 추천의 효율을 알아보고자 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서 강화학습 기법과 연관성 피드백에 대하여 알아보고 3장에서는 이 알고리즘이 신문기사를 추천하는데 어떻게 적용되는지에 대하여 살펴보고 4장에서는 실험을 통하여 제안한 방법

의 성능을 평가하고, 마지막으로 결론 및 향후 연구에 대하여 토론한다.

2. 관련연구

2.1 강화학습

기계학습을 분류하면 명시적으로 학습 목표가 주어지는 분류, 예측 등과 같은 지도기반 학습(supervised learning)과 명확한 학습의 목표가 주어지지 않는 군집화(clustering)와 같은 비지도기반 학습(unsupervised learning)으로 크게 나누어 볼 수 있다.

강화학습은 그 중간적인 특성을 띠고 있다. 이 방법은 그림 1과 같다. 학습의 주체인 에이전트는 환경의 상태를 관측하고 과거의 경험을 바탕으로 행동을 선택하고, 이에 따른 보상을 환경으로부터 받게 된다. 즉 강화학습은 동적인 환경에서 시행착오를 거치면서 환경으로부터 주어진 보상을 최대화 하기 위한 학습 방법이다 [2].

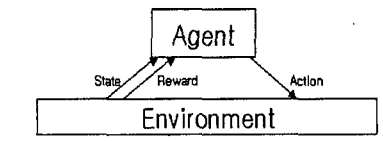


그림 1. 강화학습 framework

Q-Learning은 현존 강화학습 방법들 중 대표적인 방법으로 상태 s 에 대한 행동 a 의 가치함수 $Q(s,a)$ 를 기반으로 각 단계마다 학습자는 최대 가치함수 값을 가진 행동 a 를 선택한다. Q-Learning에서는 식 2에서 정의된 $Q(s,a)$ 를 직접 학습한다.

$$\hat{Q}(s,a) \leftarrow r + \gamma \max_{a'} \hat{Q}(s',a') \quad (1)$$

이 값은 상태 s 에서 행동 a 를 취한 후 최적으로 행동했을 경우의 보상의 총합이다. 여기서 r 은 행동 a 의 수행에 대한 보상이고, γ 는 Q 가치함수 값의 적합도를 보장하는 할인율(discount rate)이다. 먼저 행한 행동에 대한 보상 값은 높은 값으로 정하고, 같은 행동에 대해서도 시간의 흐름에 따라 그 보상 값을 감소 시킨다. 할인율 γ 는 어떤 행동이 n 번 행동에 행하여 졌을 때 보상 값은 초기 보상 값에 γ^n 을 곱한 값이 된다. Q -Learning 알고리즘은 그림 2와 같다.

1. s, a 를 초기화
2. 현재 상태 s 를 관찰
3. 행동 a 를 선택하고 실행한다.
4. 행동 a 를 행하여 보상 r 을 받는다.
5. $\hat{Q}(s,a)$ 를 다음과 같이 수정한다.

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

그림 2. Q -Learning 알고리즘

이 방법은 모델의 정보 없이 행동의 적합성을 나타내는 Q 값만을 학습하므로 구현이 간단하며 실제 여러 문제에 사용되어 좋은 결과를 보이고 있다.

2.2 연관성 피드백

연관성 피드백은 정보 추출 및 검색 분야에서 오랜 기간 연구되어 왔으며 피드백이 없을 경우에 비해 많은 성능향상을 가져오는 것으로 알려져 있다. 연관성 피드백 모델에서는 각 문서들에 대하여 사용자가 적합도 평가값을 주게 되고, 문서와 이에 해당하는 적합도 평가값으로부터 보다 평가값이 높은 문서들을 얻을 수 있도록 질의문이 수정된다[4].

3. 강화학습을 이용한 추천 시스템

기존의 연구에서는 연관성 피드백을 이용할 때 사용자가 제공한 관련 정보만을 이용하여 사용자 프로파일을 학습하는 방법을 사용하고 있다[3]. 하지만 본 연구에서는 연관성 피드백시 사용자로부터 주어지는 관련 정보와 비관련 정보를 함께 이용하여 사용자의 프로파일을 학습하는 방법을 제안한다.

여러개의 문서가 있을 때 사용자가 관심 있어 하는 문서를 추천해 주기 위하여 에이전트들은 관련 문서들을 검색하게 된다. 이때 각 문서들을 상태로 정의하고, 이 문서들을 찾는 것을 행동으로 정의한다.

각 상태에서 에이전트는 그에 따른 보상 값을 받고, 이 상태에서 관련 문서를 따라가는 행동을 취하면 이 행동이 에이전트의 최종 목적에 있어서 얼마나 유용한 것인지를 가치함수 $Q(s,a)$ 를 통해서 평가 할 수 있다.

가치함수 $Q(s,a)$ 는 문서에 대하여 사용자가 관심을 나타내는 행동에 대한 할인된 보상값의 합이 된다. 즉 사용

자의 프로파일과 문서를 나타내는 단어들을 벡터 표현으로 나타내고 사용자 프로파일상의 단어들과의 유사도를 조사하여 이 값을 보상 값으로 정의하여 가치함수 $Q(s,a)$ 를 구하였다.

상태 s 에서 행동 a 를 선택하면 가치함수 $Q(s,a)$ 로 추론을 한다. 각 상황에서 학습자는 최대 $Q(s,a)$ 값을 가진 행동을 선택한다. $Q(s,a)$ 값은 각 행동 후에 다음 학습 규칙으로 조정된다.

$$Q(s, a) \leftarrow Q(s, a) + r + w + \max_{a'} (Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)) \quad (2)$$

Q -learning과의 차이점은 즉시 보상값 r 에 비관련 정보 w 를 더하게 된다는 것이다. 다음 행동 가치함수 $Q(s,a)$ 는 모든 행동 a 에 대한 최대값을 할인하지 않는다는 것이다.

연관성 피드백은 정답이 주어지지 않고 사용자의 피드백이 주어지는 환경에서 평가를 최대화 한다는 목표를 가지므로 강화학습을 적용할 수 있다. 강화학습의 각 요소를 대응시킴으로써 피드백 방식을 구현할 수 있다.

즉 연관성 피드백을 통하여 사용자가 선호한 문서에 대한 프로파일과 사용자가 선호하지 않은 문서에 대한 프로파일을 각각 유지하게 된다. 사용자의 관련성 평가를 프로파일의 학습에 이용하게 된다. 사용자의 관련성 평가는 관련 있음(good), 관련 없음(bad), 보통(fair)의 세 가지로 평가 한다.

사용자로부터 주어진 관련 정보와 비관련 정보를 이용한 사용자 프로파일의 학습 과정은 그림 3과 같다.

1. i 문서를 처리
2. 프로파일 p 를 가지고 온다.
3. $w, Q(s,a)$ 를 초기화
4. 문서의 끝까지 반복
 s <- 현재 상태
 상태 s 에 대하여 행동 a 를 실행
 사용자의 피드백을 받은 문서에서 용어를 추출
 보상 값을 준다.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + r + w + \max_{a'} (Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))$$
5. 프로파일 업데이트
6. $i=i+1$
 goto step 1

그림 3. learning pseudo code

학습된 결과를 이용하여 에이전트가 사용자의 관심도에 나타난 단어들의 Q 값이 가장 큰 문서를 추천하게 됨으로써 효율적인 추천을 할 수 있다. Q -Learning은 미래의 보상 값을 할인하기 때문에 장기간 동안 보상 값을 유지하는 행동보다 단기간 내 끝나는 행동을 선호한다.

신문기사를 추천해주는 시스템은 사용자로부터 피드백을 받아 프로파일을 구성하고, 추천 시스템은 사용자의 프로파일을 이용하여 사용자가 선호할 만한 문서를 추천해 주게 된다. 추천 시스템의 구성은 그림 4와 같다.

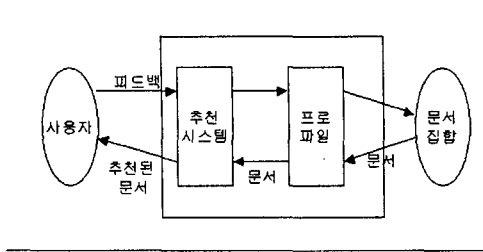


그림 4. 추천 시스템 구성

사용자의 관심도에 나타난 단어들의 Q 값 이용하여 상위 랭크된 문서를 추천함으로써 더 효율적인 추천을 할 수 있다. 이는 특정 도메인이라는 한정된 상황에서 성능에 많은 향상을 가져올 수 있다. 그러나 실제로 사용자의 프로파일은 사용자의 관심도 변화에 따라서 달라질 수 있다. 본 논문에서 제안한 추천 알고리즘은 그림 5와 같다.

1. 사용자 프로파일 생성
2. 문서의 벡터 생성
3. 프로파일과 문서의 유사도 계산
4. 가치함수 Q 계산
5. 단계 2로 가서 반복
6. 만족된 결과 있으면 리턴

그림 5. 추천 알고리즘 의사코드

4. 실험 및 결과

실험에 사용된 데이터로는 중앙일보 신문 기사를 대상으로 수집된 총 400개의 문서를 사용한다. 이 문서의 구성은 중앙일보 웹 사이트에서 제공되는 10개의 카테고리에서 각각 40개의 문서를 수집 사용하였고, 문서의 내용은 정치, 경제, 스포츠 등으로 구성되어 있다.

추천 시스템의 성능을 평가하기 위해 3개의 토픽에 대하여 관련성 평가를 받는다. 3개의 토픽은 아데네 올림픽, 경제 기사 중 유가 관련 기사, 문화연예 관련 기사 중 연예계관련 기사로 구성되어진다. 실험 조건상 사용자 관심도 변화는 없는 것으로 가정하였다.

문서들을 전처리 과정을 통하여 TF벡터로 표현하고 이 문서들에 나오는 단어들과 사용자의 프로파일에 나오는 단어들을 이용하여 보상함수 $R(x)$ 를 구하고, 사용자 프로파일과 문서간의 유사도를 측정하기 위한 유사도 정량화는 두 벡터 간의 코사인 유사도 측정법을 사용한다.

$$sim(d_j, p) = \frac{d_j \cdot p}{|d_j| \cdot |p|} = \frac{\sum_{i=1}^n w_{i,j} \times w_{i,p}}{\sqrt{\sum_{i=1}^n w_{i,j}^2} \times \sqrt{\sum_{i=1}^n w_{i,p}^2}} \quad (3)$$

위의 식에서 p 는 사용자 프로파일 벡터를 나타내고, d 는 문서의 벡터로 표현된다.

실험은 각 토픽별로 관련 정보를 이용하였을 경우와 비관련 정보도 함께 이용하였을 경우의 정확률과 재현율을 이용하여 평가 하였다.

추천 시스템에서 제공되는 상위 문서 10, 20개의 문서를 대상으로 한다. 제안된 방법을 추천 시스템에 적용한 실험결과는 다음 표 1, 2와 같다.

표 1. 추천 시스템의 정확률

	relevant			relevant & irrelevant		
	topic1	topic2	topic3	topic1	topic2	topic3
10	60	70	50	65	73	54
20	45	78	60	70	85	67

표 2. 추천 시스템의 재현율

	relevant			relevant & irrelevant		
	topic1	topic2	topic3	topic1	topic2	topic3
10	50	52	52	70	67	64
20	80	65	67	73	82	75

표 1에서 볼 수 있듯이 사용자로부터 받은 관련성 정보만을 이용한 경우 보다 비관련성 정보도 같이 이용한 방법이 더 나은 추천 성능을 보이는 것을 알 수 있다.

신문기사 추천 시스템에서는 특정 토픽에 대한 추천이 일반적인 토픽보다 더 나은 성능을 보이고 있다.

5. 결론

본 연구에서는 강화학습 기법과 관련성 정보를 이용하여 사용자 프로파일을 학습하고, 이를 신문 기사 추천 시스템에 적용하였다. 본 연구에서 제안한 방법은 특정 도메인에서 사용자가 원하는 정보를 추천하는데 있어서 크게 도움이 될 것이다. 향후 연구에서는 이번 연구를 바탕으로 사용자의 관심도를 학습하는데 중점을 두고 문서를 추천하는 방법을 생각해 볼 수 있다.

참고문헌

- [1] Seo, Y., Zang, B., " *Personalized Web Document Filtering Using Reinforcement Learning*," Applied Artificial Intelligence, Vol. 15, 2001.
- [2] Richard S. Sutton, Andrew G. Barto, " *Reinforcement Learning: An Introduction*," MIT Press, 1998.
- [3] Watkins, C. and Dayan, P., " *Q-Learning*," Machine Learning, 8(3):279-292, 1992.
- [4] Leski, A., " *Relevance and Reinforcement in Interactive Browsing*," International Conference on Information and Knowledge Management, pp. 119-126, 2000.
- [5] Tom M. Mitchell, *Machine Learning*, McGraw-Hill com. Inc., 1997.