

## 유사도와 난이도를 이용한 학습 콘텐츠 추천 방법

박재욱\* 이용규\*

### A Method for Recommending Learning Contents Using Similarity and Difficulty

Jae Wook Park\* Yong Kyu Lee\*

#### 요약

이러닝 시스템에서 학습자에게 적합한 콘텐츠 선택을 돕기 위한 콘텐츠 추천 시스템은 필수적이다. 학습자의 선호도를 통한 콘텐츠 추천은 협업 필터링 추천 방법과 내용 기반 추천 방법이 가장 많이 사용되고 있다. 그러나 기존 추천 방법들은 학습자의 학습수준을 고려하지 않고 다른 사용자의 선호도를 기반으로 학습 콘텐츠를 추천한다. 따라서 상대적으로 콘텐츠를 학습한 학습자가 적은 경우 추천의 효율성이 떨어지고, 새로운 아이템이 추가될 경우 추천이 쉽지 않은 단점이 있다. 이 문제를 해결하기 위해 우리는 학습 콘텐츠의 유사도와 난이도에 기반한 콘텐츠 추천 방법을 제안한다. 학습 콘텐츠의 두 특성을 반영한 추천함수에 의해 선행학습 성취도가 낮은 학습자에게는 난이도가 낮고 유사도가 높은 콘텐츠를 추천하고, 성취도가 높은 학습자에게는 난이도가 높고 유사도가 낮은 콘텐츠를 추천한다. 이와 같이 다른 학습자의 선호도와는 무관하게 학습자의 성취도에 따라 가장 적합한 콘텐츠를 추천할 수 있다.

▶ Keyword : 학습 콘텐츠, 추천 시스템, 유사도, 난이도

#### Abstract

It is required that an e-learning system has a content recommendation component which helps a learner choose an item. In order to predict items concerning learner's interest, collaborative filtering and content-based filtering methods have been most widely used. The methods recommend items for a learner based on other learner's interests without considering the knowledge level of the learner. So, the effectiveness of the recommendation can be reduced when the number of overall users are relatively small. Also, it is not easy to recommend a newly added item. In order to address the problem, we propose a content recommendation method based on the similarity and the difficulty of an item. By using a recommendation function that reflects both characteristics of

• 제1저자 : 박재욱 • 교신저자 : 이용규

• 투고일 : 2011.06.01, 심사일 : 2011.06.28, 게재확정일 : 2011.07.12.

\* 동국대학교 컴퓨터공학과-서울(Dept. of Computer Science and Engineering, Dongguk University-Seoul)

items, a higher-level learner can choose more difficult but less similar items, while a lower-level learner can select less difficult but more similar items, Thus, a learner can be presented items according to his or her level of achievement, which is irrelevant to other learner's interest.

▶ Keyword : Learning Content, Recommendation System, Similarity, Difficulty

## I. 서 론

인터넷의 발달 및 확산과 더불어 정보를 효과적으로 검색하기 위해 추천 시스템에 대한 많은 연구가 있었다 [1][4][5][6][7]. 이러닝 시스템에서도 학습자 수준에 맞는 적절한 학습 콘텐츠를 추천하기 위한 방법은 맞춤형 학습, 자기주도 학습시스템에 있어 매우 중요한 요소이다.

기존의 학습 콘텐츠 추천 방법들은 크게 내용 기반 추천 방법[7][14]과 협업 필터링 추천 방법[8][16][18][19]으로 나뉘고, 그밖에 온톨로지를 이용한 규칙 기반 추천 방법 [9][10][11] 및 최근에는 추천 방법을 혼용한 하이브리드 추천 방법[2][12][13][15] 등이 제안되었다.

그러나 기존 방법들은 난이도나 성취도 등을 고려하지 않고 콘텐츠의 유사도나 학습자의 선호도 등을 고려하여 추천하기 때문에 학습자의 수준보다 너무 어렵거나 또는 너무 쉬운 콘텐츠를 추천할 수 있는 단점이 있고, 학습자가 학습한 콘텐츠가 적거나 전체 학습자가 너무 적을 경우 정확한 추천을 할 수 없는 단점이 있다.

본 논문에서는 이러한 단점을 보완하기 위하여 콘텐츠의 난이도, 학습자의 선행학습에 관한 성취도, 콘텐츠 간의 유사도 등을 계산하여 이들간의 상관관계에 따라 후행학습을 추천하는 방법을 제안한다. 이를 통해 학습자 수준에 적합한 콘텐츠를 객관적으로 추천할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 학습 콘텐츠 추천 방법에 대한 기존 연구에 대하여 기술하고, 3장에서는 난이도, 유사도, 성취도 등을 고려한 추천 방법에 대하여 기술하고, 4장에서는 시스템 구성에 관하여 기술하고, 5장에서는 제안한 추천 방법에 대한 객관성을 검증하기 위해 초등학교 영어 시험인 JET[20]의 문제지를 이용하여 실험한 결과를 기술하고, 6장에서는 결론을 맺는다.

## II. 관련 연구

이러닝 시스템에서 높은 효율을 제공하기 위해 학습자의 수준, 취향, 동기 등을 고려하여 개인 맞춤형 시스템에서 적

절한 학습아이템을 추천하는 방법은 매우 중요하다.

추천 방법에 관한 연구는 크게 내용 기반 추천 방법 [7][14]과 협업 필터링 추천 방법[8][16][18][19]으로 나뉘어 연구되어왔다.

내용 기반 추천 방법은 학습자의 선호도 프로파일의 속성과 아이템 속성간의 유사도를 측정하여 유사도가 높은 경우 즉, 예측 선호도가 높은 콘텐츠를 추천하는 방법으로 텍스트 기반의 추천 시스템에서 많이 이용되었는데 사용자 선호도 프로파일은 과거에 학습하였던 아이템으로부터 속성 값을 계산한다.

Chen[7]은 콘텐츠의 특징을 표현할 수 있는 분류, 스타일, 학습시간 등의 속성들로 콘텐츠의 벡터를 만들고, 과거에 학습 했던 아이템들의 선호도 프로파일을 벡터로 만든 후 각 속성별로 가중치를 두어 이들의 곱의 합이 높은 순으로 콘텐츠를 추천하는 방법을 제안하였고, Ghauth[14]는 벡터스페이스 모델을 이용하여 콘텐츠 내용간의 유사도를 계산하고, 다른 학습자들의 콘텐츠에 대한 평점의 평균을 예측 선호도로 계산하여 유사도가 높은 콘텐츠 중 평점의 평균이 높은 콘텐츠를 추천하는 방법을 제안하였다.

그러나 내용 기반 추천 방법은 학습자의 수준이나 해당 콘텐츠의 난이도는 고려하지 않고 예측 선호도가 높은 콘텐츠만 추천하며, 학습자가 학습한 콘텐츠가 적을 경우 정확히 선호도를 예측하기 어려운 단점이 있다.

협업 필터링 추천 방법은 학습자와 취향이 비슷한 학습자 그룹을 찾고 그 학습자 그룹이 선호하는 학습 콘텐츠를 추천하는 방법이다. 내용 기반 추천 방법과의 차이점은 아이템의 내용을 고려하지 않고 학습자들의 기호에만 의존한다는 것이다.

Resnick[16]이 발표한 GroupLens가 대표적인 협업 필터링 추천 방법으로 사용자들이 본 뉴스에 대해 명시적 평점을 부여하고 사용자간의 상관계수를 이용하여 특정 사용자의 보지 않은 뉴스에 대한 선호도를 예측하여 추천하는 방법을 제안하였다.

Tan[8]은 피어슨 상관계수를 이용하여 학습자들이 평정한 평점으로 학습자간의 유사도를 계산하고, 유사도가 높은 근접한 이웃들로부터 연관규칙에 기반하여 top-N개의 학습아이템을 추천하는 이러닝 시스템을 제안하였다.

그러나 협업 필터링 추천 방법은 새로운 콘텐츠가 추가되었을 때 다른 학습자의 콘텐츠에 대한 선호도가 없기 때문에

신규 아이টে을 추천할 방법이 없고, 학습자가 적으면 콘텐츠가 많아도 추천할 콘텐츠의 범위가 줄어드는 단점이 있다.

본 논문에서는 이러한 단점을 보완하기위해 학습자의 성취도와 선행학습과의 유사도, 각 콘텐츠의 난이도를 계산하여 보다 객관적인 콘텐츠를 추천할 수 있는 방법을 제안한다.

### III. 난이도와 유사도 기반 후행학습 추천 방법

#### 1. 콘텐츠 추천 프로세스

그림 1은 후행학습 콘텐츠를 추천하는 프로세스이다. 선행학습이 끝나면 학습에 대한 테스트를 시행하여 성취도를 계산한다. 그리고 다른 콘텐츠간의 유사도 및 난이도를 계산한다.

성취도가 높을 경우 선행학습에 대한 이해도가 높다고 볼 수 있으므로 선행학습과 비슷한 콘텐츠나 쉬운 콘텐츠를 추천할 경우 비효율적이고 학습자가 흥미를 잃을 수 있기 때문에 학습범위의 확대와 학습수준을 높이기 위하여 선행학습보다는 난이도가 높고 유사도가 낮은 콘텐츠를 추천한다. 성취도가 낮은 경우에는 선행학습에 대한 이해도가 낮다고 볼 수 있으므로 학습효과를 높이기 위하여 선행학습과 유사한 내용의 콘텐츠 중 선행학습보다 쉬운 콘텐츠를 추천한다.

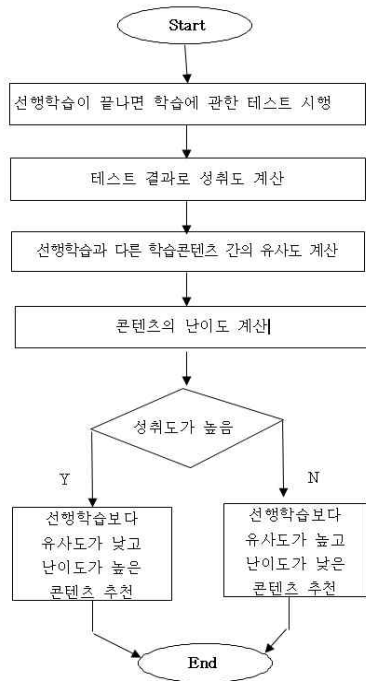


그림 1. 콘텐츠 추천 프로세스  
Fig. 1. Content recommendation Process

#### 2. 성취도 계산

성취도는 선행학습이 종료된 후 테스트를 거쳐 테스트 성적을 백분율로 표시하여 0과 1사이의 값을 갖도록 하며, 1에 가까울수록 성취도가 높다.

#### 3. 유사도 계산

유사도는 벡터간의 방향성의 차이 즉, 코사인 측정법으로 유사도를 계산하는 벡터스페이스 모델(Vector space model) 또는 벡터간의 거리로 유사도를 계산하는 유클리디언 거리(Euclidean Distance)로 계산한다[17]. 벡터는 모든 콘텐츠에서 사용된 단어의 집합으로 구성된 색인어를 이용하여 각 콘텐츠 별 색인어의 빈도수로 표현한다.

##### 3.1 벡터스페이스 모델

벡터스페이스 모델은 벡터간의 방향성의 차이 즉, 코사인 측정법으로 표현할 수 있다[17]. 따라서 유사도는 0과 1사이의 값을 갖게 되며 1에 가까울수록 유사도가 높다. 벡터간의 각도가 적으면 유사도가 높고 각도가 크면 유사도가 낮다.

벡터스페이스 모델은 다음과 같은 수식으로 표현한다.

$$\cos(Q, D) = \frac{Q \cdot D}{|Q||D|} = \frac{\sum_{t=1}^n q_t \cdot d_t}{\sqrt{\sum_{t=1}^n (q_t)^2} \sqrt{\sum_{t=1}^n (d_t)^2}} \dots\dots [수식 1]$$

- Q : 선행학습 콘텐츠 벡터
- D : 선행학습 콘텐츠와 비교할 학습 콘텐츠
- t : 색인어
- qt : 선행학습 콘텐츠 벡터(Q)에서 색인어 t에 대한 가중치
- dt : 선행학습 콘텐츠 벡터(Q)와 비교할 벡터(D)의 색인어 t에 대한 가중치

[수식 1]에서 선행학습 콘텐츠(Q)의 색인어 가중치 qt와 선행학습과 비교할 콘텐츠(D)의 색인어 가중치 dt는 TF(Term Frequency)\*IDF(Inverse Document Frequency) 방식을 적용하여 다음과 같이 계산된다.

$$v_t = tf \times idf = f_t \cdot \log(N/F_t) \dots\dots\dots [수식 2]$$

- vt : 콘텐츠벡터 v 에서의 색인어 t의 가중치 ([수식1]에서 qt와 dt)
- tf : 벡터 v에서의 색인어 t의 빈도수 (Term Frequency)

- idf : 벡터 v에서의 색인어 t의 역 빈도수 (Inverse Document Frequency)
- ft : 콘텐츠 벡터 v에서의 색인어 t의 빈도수
- Ft : 단어 t를 포함하는 콘텐츠 총 개수
- N : 콘텐츠의 총 개수

[수식 2]에서 IDF는 단어 t에 대한 콘텐츠의 가중치로서 빈도수가 적게 나오는 단어가 중요한 정보를 지닌다고 보고  $\log(N/f_t)$ 로 계산한다.

### 3.2 유클리디언 거리

콘텐츠의 상대적 크기가 비슷할 경우 유클리디언 거리를 이용하여 유사도를 계산 한다[17]. 유클리디언 거리를 이용하여 유사도를 계산할 때는 [수식 3]으로 표현할 수 있다. 유클리디언 거리를 계산하여 값이 작으면 즉, 거리가 가까우면 유사도가 높고, 값이 크면 즉, 거리가 멀수록 유사도가 낮다.

$$dist(Q, D) = \sqrt{\sum_{i=1}^n (q_i - d_i)^2} \dots\dots\dots [수식 3]$$

- Q : 선행학습 콘텐츠 벡터
- D : 선행학습 콘텐츠와 비교할 학습 콘텐츠 벡터
- t : 색인어
- qt : 선행학습 콘텐츠 벡터(Q)에서 색인어 t에 대한 빈도수
- dt : 선행학습 콘텐츠 벡터(Q)와 비교할 벡터(D)의 색인어 t에 대한 빈도수

유클리디언 거리를 계산할 때 색인어의 별도의 가중치를 고려하지 않고 빈도수만 계산한다. [수식 3]을 0과 1사이의 값이 나오도록 정규화 하여 [수식 4]와 같이 표현할 수 있다. [수식 4]의 값이 1에 가까울수록 유사도가 높다.

$$dist(Q, D) = \frac{1}{1 + \sqrt{\sum_{i=1}^n (q_i - d_i)^2}} \dots\dots\dots [수식 4]$$

- Q : 선행학습 콘텐츠 벡터
- D : 선행학습 콘텐츠와 비교할 학습 콘텐츠 벡터
- t : 색인어
- qt : 선행학습 콘텐츠 벡터(Q)에서 색인어 t에 대한 빈도수
- dt : 선행학습 콘텐츠 벡터(Q)와 비교할 벡터(D)의 색인어 t에 대한 빈도수

### 4. 난이도 계산

각 콘텐츠의 난이도는 콘텐츠 제작자 또는 콘텐츠 전문가가 부여할 수 있고, 용어들의 난이도를 정의한 용어사전을 활용하거나 온톨로지에서 용어들의 심화관계를 이용하여 콘텐츠의 색인단어들에 의해 각 콘텐츠의 난이도를 계산할 수 있다[3].

온톨로지를 이용하여 난이도를 계산할 경우, 온톨로지의 개체는 학습 콘텐츠에서 사용된 색인단어들로 구성한다. 단어 간의 관계는 상하위 관계만 고려하며, 단어의 심화도를 속성으로 부여한다. 온톨로지는 상위단어가 두 개 이상일 수 있는 그래프 구조이므로 이를 트리구조로 변경하여 난이도를 계산한다. 상위단어가 두 개 이상인 n개의 구조로 되어 있는 단어는 해당단어의 하위 단어들을 n-1번 복사하여 상위단어의 서브트리로 만들어 트리구조로 변경한다.

각 콘텐츠의 난이도는 콘텐츠에서 사용된 색인단어들을 온톨로지에서 찾고 사용된 단어들 간의 심화도를 거리로 표현하여 거리의 합으로 난이도를 계산한다. 온톨로지에서 부모와 자식 노드간의 거리는 1로 설정하며, 거리가 길수록 난이도가 높은 단어이다, 콘텐츠에 사용된 단어들의 거리의 합을 정규화하여 0과 1사이의 난이도를 가지며 1에 가까울수록 난이도가 높아지도록 [수식 5]와 같이 표현한다.

$$Diff(c) = \frac{\sum_{i=1}^n d(t_i) \times f(t_i)}{(\sum_{i=1}^n f(t_i)) \times D} \dots\dots\dots [수식 5]$$

- Diff(c): 콘텐츠 c의 난이도
- n : 색인단어의 개수
- d(ti) : 루트에서부터 단어 ti 까지의 거리의 합
- f(ti) : 단어 ti의 빈도수
- D: 온톨로지의 단어의 거리 중 가장 큰 값

### 5. 후행 학습 추천

성취도, 유사도, 난이도가 정해지면 다음과 같이 성취도, 유사도, 난이도 간의 상관관계를 함수로 표현하여 성취도가 높은 경우에는 난이도가 높고 유사도가 낮은 콘텐츠를 추천하고, 성취도가 낮은 경우에는 난이도가 낮고 유사도가 높은 콘텐츠를 추천한다.

#### 5.1 성취도와 유사도의 상관관계 함수

성취도와 유사도의 상관관계 함수는 다음과 같이 [수식 6]으로 표현한다.

$$C_{g,s} = 1 - |g - (1 - s)| \dots\dots\dots [수식 6]$$

- C<sub>g,s</sub> : 성취도와 유사도의 상관관계 함수  
(0 ≤ C<sub>g,s</sub> ≤ 1)
- g : 성취도(0 ≤ g ≤ 1)
- s : 유사도(0 ≤ s ≤ 1)

[수식 6]에서 성취도(g)와 유사도(s)의 함수 값은 두 값의 차이가 커지면 증가하고 두 값의 차이가 적으면 감소한다. 성취도가 높으면 유사도가 낮은 경우 함수 값이 크게 나온다. 반대로 성취도가 낮으면 유사도가 높은 경우 함수 값이 크게 나온다. 즉, 위 함수에서 성취도와 유사도는 반비례 관계이다.

5.2. 성취도와 난이도의 상관관계 함수

성취도와 난이도의 상관관계 함수는 다음과 같이 [수식 7]로 표현한다.

$$C_{g,d} = 1 - |g - d| \dots\dots\dots [수식 7]$$

- C<sub>g,d</sub> : 성취도와 난이도의 상관관계 함수  
(0 ≤ C<sub>g,d</sub> ≤ 1)
- g : 성취도(0 ≤ g ≤ 1)
- d : 난이도(0 ≤ d ≤ 1)

[수식 7]에서 난이도(d)와 성취도(g)의 함수 값은 두 값의 차이가 적으면 증가하고, 두 값의 차이가 커지면 감소한다. 성취도가 높으면 난이도가 높은 경우 함수 값이 크게 나온다. 반대로 성취도가 낮으면 난이도가 낮은 경우 함수 값이 크게 나온다. 즉, 위 함수에서 난이도와 성취도는 비례 관계이다.

5.3 추천함수

[수식 6]과 [수식 7]에 각각 가중치를 부여하여 [수식 8]과 같이 콘텐츠를 추천하는 복합함수를 만들어 함수 값이 높은 콘텐츠를 추천한다.

$$C_{g,s,d} = w_1(1 - |g - (1 - s)|) + w_2(1 - |g - d|) \dots\dots [수식 8]$$

- C<sub>g,s,d</sub> : 성취도와 유사도와 난이도간의 상관관계를 이용한 콘텐츠 추천함수 (0 ≤ C<sub>g,s,d</sub> ≤ 1)
- g : 성취도
- d : 난이도
- s : 유사도
- w<sub>1</sub> : 유사도에 대한 가중치
- w<sub>2</sub> : 난이도에 대한 가중치

[수식 8]에서 성취도(g)가 높은 경우에는 난이도(d)가 높고 유사도(s)가 낮은 경우 값이 높게 나타난다. 반대로, 성취도(g)가 낮은 경우에는 난이도(d)가 낮고 유사도(s)가 높은 경우 값이 높게 나타난다. 즉, 1에 가까운 값을 갖는 콘텐츠를 추천한다.

IV. 시스템구성

그림 2는 추천 시스템의 구성도이다. 데이터베이스에서 온톨로지는 모든 콘텐츠의 색인단어가 온톨로지 형태로 저장되어 있고, 콘텐츠에는 모든 콘텐츠의 학습 내용이 저장되어 있으며, 학습관리에는 로그인 정보, 학사관리 정보 등이 저장되어 있다. 학습자는 학습운영 프로그램을 통해 학습을 하고 학습이 종료되면 테스트를 거쳐 성취도를 계산한다. 그리고 유사도 계산모듈에서 선행학습과 콘텐츠간의 유사도를 계산하며, 콘텐츠 색인어 온톨로지를 이용해 난이도 계산모듈에서 난이도를 계산한다. 후행학습 추천모듈에서 추천함수를 이용해 함수 값이 높은 학습 콘텐츠를 추천한다.

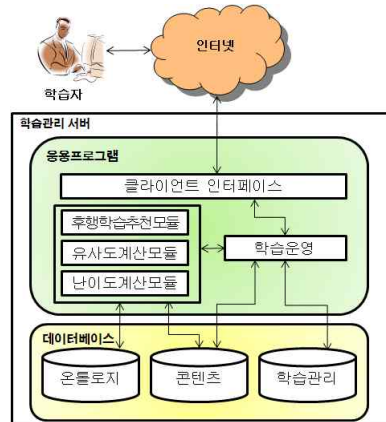


그림 2 추천 시스템 구성  
Fig. 2. Recommendation System Organization

V. 실험 및 분석

실험은 추천함수의 적정성 여부를 검증하기 위해 YBM Si-sa에서 초등학교 학생들의 영어능력평가를 위해 개발한 JET(Junior English Test)[20]의 기출문제를 대상으로 하였다. JET는 듣기영역과 쓰기영역으로 구성되어 있으며 초급, 중급, 고급의 3등급으로 되어있다. 이 중 각 등급별 4개의 문제지를 대상으로 하여 각 문제지의 쓰기영역 문제만 선

택하여 실험을 실시하였다.

표 1은 '초급1', '중급1', '고급1' 문제지로 선행학습을 했을 경우 다른 문제지와 유사도 및 각 문제지의 난이도 계산결과이다.

표 1. JET 문제지의 난이도 및 문제지 간의 유사도  
Table 1. The difficulty of JET Tests and the Similarity between Tests

문제지	난이도 (d)	초급1과의 유사도 (s1)	중급1과의 유사도 (s2)	고급1과의 유사도 (s3)
초급1	0.40	1.0	0.02	0.03
초급2	0.39	0.09	0.01	0.02
초급3	0.37	0.16	0.02	0.01
초급4	0.40	0.05	0.03	0.03
중급1	0.50	0.02	1.0	0.05
중급2	0.48	0.02	0.06	0.07
중급3	0.43	0.07	0.09	0.08
중급4	0.43	0.03	0.04	0.07
고급1	0.57	0.03	0.05	1.0
고급2	0.53	0.04	0.06	0.09
고급3	0.50	0.01	0.06	0.07
고급4	0.52	0.01	0.05	0.04

난이도는 12개의 문제지에 나온 610개의 명사와 동사 단어를 상하위 구조의 온톨로지로 구성하고, 각 단어별 심화도를 부여한 후 문제지에 사용된 단어들의 심화도의 합으로 계산하였고[3], 유사도는 벡터스페이스 모델을 이용하여 계산하였다.

표 2는 선행학습의 시험성적 즉, 성취도의 높고 낮음에 따라 어떤 콘텐츠를 추천하는지를 측정하기 위하여 성취도가 낮은 경우를 0.3, 중간일 경우를 0.5, 높은 경우를 0.9라고 임의로 부여하고 [수식 8]의 추천함수 값을 계산한 결과이다. 가중치 w1과 w2는 각각 0.5로 계산하였으며, 표 2의 (a), (b), (c)는 선행학습이 각각 초급1, 중급1, 고급1인 경우의 추천함수 값이다.

성취도가 낮은 경우는 학습한 콘텐츠의 내용에 대한 이해도가 떨어진다고 볼 수 있으므로 유사한 내용의 콘텐츠를 재추천하면서 학습효과를 높이기 위하여 선행학습보다 쉬운 콘텐츠가 추천되어야한다.

표 2. 추천함수 값  
Table 2. The Value of Recommendation functions  
(a) 선행학습이 초급1일 경우

문제지	추천함수 값		
	성취도 (g=0.3)	성취도 (g=0.5)	성취도 (g=0.9)
초급2	0.65	0.74	0.74
초급3	0.70	0.765	0.71
초급4	0.63	0.725	0.73
중급1	0.56	0.76	0.76
중급2	0.57	0.75	0.75
중급3	0.62	0.75	0.75
중급4	0.60	0.73	0.73
고급1	0.53	0.73	0.80
고급2	0.56	0.755	0.79
고급3	0.56	0.755	0.76
고급4	0.55	0.745	0.77

(b) 선행학습이 중급1일 경우

문제지	추천함수 값		
	성취도 (g=0.3)	성취도 (g=0.5)	성취도 (g=0.9)
초급1	0.61	0.71	0.71
초급2	0.61	0.70	0.70
초급3	0.63	0.70	0.70
초급4	0.62	0.72	0.72
중급2	0.59	0.77	0.77
중급3	0.63	0.76	0.76
중급4	0.61	0.74	0.74
고급1	0.54	0.74	0.81
고급2	0.57	0.77	0.80
고급3	0.58	0.78	0.78
고급4	0.57	0.77	0.79

(c) 선행학습이 고급1일 경우

문제지	추천함수 값		
	성취도 (g=0.3)	성취도 (g=0.5)	성취도 (g=0.9)
초급1	0.62	0.72	0.72
초급2	0.62	0.71	0.71
초급3	0.62	0.69	0.69
초급4	0.62	0.72	0.72
중급1	0.58	0.78	0.78
중급2	0.60	0.78	0.78
중급3	0.63	0.76	0.76
중급4	0.62	0.75	0.75
고급2	0.58	0.78	0.81
고급3	0.59	0.79	0.79
고급4	0.56	0.76	0.78

또한, 성취도가 높은 경우는 콘텐츠 내용에 대한 이해도가 높다고 볼 수 있으므로 내용이 유사한 콘텐츠나 쉬운 콘텐츠를 추천하면 흥미를 잃을 수 있기 때문에 선행학습 보다 난이도가 높고 유사성이 낮은 콘텐츠를 추천해야 한다.

실험결과 성취도가 높은 경우에는 난이도가 높고 유사도가 낮은 문제지의 추천합수 값이 높았으며, 성취도가 낮은 경우에는 난이도가 낮고 유사도가 높은 문제지의 추천합수 값이 높았다.

기존 추천 방법으로 추천을 했을 경우 유사도나 타 학습자의 선호도만을 고려하기 때문에 학습자 수준과는 맞지 않는 콘텐츠가 추천될 수 있고 신규 콘텐츠가 추가 되었을 경우 학습자들의 선호도 프로파일이 없기 때문에 추천되기 힘들다.

실험결과를 통하여 제안한 방법이 내용의 유사도뿐만 아니라 학습자 수준에 따라 적합한 콘텐츠가 추천됨을 알 수 있었다.

표 3은 제안방법과 기존 추천 방법과의 차이점을 요약한 표이다. 제안한 방법은 학습자 수준을 고려하여 추천하기 때문에 학습자 능력에 맞는 콘텐츠를 객관적으로 추천할 수 있고, 기존 추천 방법의 선호도 예측 및 타 학습자의 평가 등이 필요 없기 때문에 신규 콘텐츠가 추가되더라도 누락되지 않고 추천될 수 있는 장점이 있다.

표 3. 제안방법과 기존 추천 방법의 차이점  
Table 3. Comparison between the Proposed Method and the Previous Methods

구분	제안방법	내용 기반 추천 방법	협업 필터링 추천 방법
학습자 수준	고려	고려안함	고려안함
신규콘텐츠 추천	가능	가능	불가능
타학습자 평가	필요없음	필요없음	필요
선호도 예측	없음	필요	필요

## VI. 결 론

기존 학습 콘텐츠 추천 방법은 내용의 유사성이나 학습자의 선호도만을 고려하기 때문에 학습자의 수준에 맞는 콘텐츠를 추천하는데 한계가 있으며, 신규 콘텐츠가 발생하거나 학습자가 적으면 선호도를 예측하기가 쉽지 않은 단점이 있었다.

본 논문에서는 이러한 단점을 해결하기 위해 학습자의 성취도, 콘텐츠의 난이도, 콘텐츠간의 유사도 등을 객관적인 방법으로 계산하고 이들의 상관관계를 통하여 적합한 콘텐츠를 추천하는 방법을 제안하였다.

성취도가 낮은 학습자에게는 선행학습의 이해도를 높이기 위해 비슷한 내용의 콘텐츠와 쉬운 콘텐츠를 추천해야 하므로 난이도가 낮고 유사도가 높은 콘텐츠를 추천하며, 성취도나 높은 학습자에게는 흥미를 유발시키고 실력향상을 위하여 학습의 범위를 넓히고 선행학습보다 어려운 콘텐츠를 추천해야 하므로 난이도가 높고 유사도가 낮은 콘텐츠를 추천한다.

콘텐츠의 유사도는 벡터스페이스 모델 또는 유클리디언 거리를 이용하여 계산하고 난이도는 콘텐츠 제작자가 지정하거나, 콘텐츠에 사용된 단어들의 상하위 구조로 구성된 온톨로지를 이용하여 계산할 수 있다.

제안한 방법의 적정성을 검증하기 위해 초등학교 영어시험인 JET 문제지의 난이도 및 유사도를 계산하여 실험한 결과 추천합수에 의해 학습자 수준에 적절한 콘텐츠를 추천함으로써 학습자 수준과는 무관하게 콘텐츠에 대한 평가나 선호도만을 고려하는 기존 추천 방법들의 단점을 해결하였다. 또한, 기존 방법들은 신규 콘텐츠가 추가될 경우 다른 학습자의 평가나 선호도가 없으므로 신규 콘텐츠가 추천되기 어려우나, 제안한 방법은 신규 콘텐츠가 추가되더라도 다른 학습자의 평가나 선호도가 필요 없기 때문에 신규 콘텐츠를 포함하여 가장 적합한 콘텐츠를 추천할 수 있는 장점이 있다.

본 논문에서 제안한 방법은 텍스트 기반의 콘텐츠 추천 위주로 다루었지만, 멀티미디어 환경의 이러닝 콘텐츠를 추천하는데도 활용이 가능하다.

## 참고문헌

- [1] Byeong Man Kim, Qing Li, Si Gwan Kim, En Ki Lim, Ju Yeon Kim, "A New Approach Combining Content-based Filtering and Collaborative Filtering for Recommender Systems," Journal of The Korean Institute of Information Scientists and Engineers, Vol. 31, No. 3, pp. 332-342, Mar. 2004.
- [2] Yong Kim, Sung Been Moon, "A Study on Hybrid Recommendation System Based on Usage frequency for Multimedia Contents," Journal of The Korea Society for Information management, Vol. 23, No. 3, pp. 91-125, Sep. 2006.
- [3] Jae Wook Park, Mee Hwa Park, Yong Kyu Lee, "An Ontology-Based Method for Calculating the Difficulty of a Learning Content," Journal of The

- Korea Society of Computer and Information, Vol. 16, No. 2, pp. 83-91, Feb. 2011.
- [4] Mojdeh Talabeigi, Rana Forsati, Mohammad Reza Meybodi, "A Dynamic Web Recommender System Based on Cellular Learning Automata," Proceedings of the IEEE International Conference on Computer Engineering and Technology, pp. 755-761, Apr. 2010.
- [5] Jian Chen, Roman Y. Shtykh, Qun Jin, "A Web Recommender System Based on Dynamic Sampling of User Information Access Behaviors," Proceedings of the IEEE International Conference on Computer and Information Technology, pp. 172-177, Oct. 2009.
- [6] Yi-Hung Wu, Yong-Chuan Chen, Arbee L. P. Chen, "Enabling Personalized Recommendation on The Web Based on User Interests and Behaviors," Proceedings of IEEE International Workshop on Research Issues in Data Engineering, pp. 17-24, Apr. 2001.
- [7] Hong-Ren Chen, "Learning Object Recommendation Services in Interactive E-Learning Systems," Proceedings of the 5th WSEAS International Conference on E-ACTIVITIES, pp. 410-414, Nov. 2006.
- [8] Huiyi Tan, Junfei Guo, Yong Li, "E-Learning Recommendation System," Proceedings of IEEE International Conference on Computer Science and Software Engineering, pp. 430-433, Apr. 2008.
- [9] Li-ping Shen, "Ontology-based Learning Content Recommendation," International Journal of Continuing Engineering Education and Life Long Learning, Vol. 15, No. 3, pp. 308-317, June 2005.
- [10] Feng-Jung Liu, "Design of Self-directed E-Learning Material Recommendation System with On-line Evaluation," Proceedings of IEEE International Conference on Hybrid Information Technology, pp. 274-277, Aug. 2008.
- [11] Saman Shishechi, Seyed Yashar Banihasem, Nor Azan Mat Zin, "A Proposed Semantic Recommendation System for E-Learning," Proceedings of IEEE International Symposium on Information technology, pp. 1-5, June 2010.
- [12] Ehsan Emadzadeh, Amade Nikfarjam, Kairil Imran Ghauth, Ng Kok Why, "Learning Materials Recommendation Using a Hybrid Recommender System with Automated Keyword Extraction," World Applied Science Journal, Vol. 9, No. 11, pp. 1260-1271, 2010.
- [13] Leyla Zhuhadar, Olfa Nasraoui, Robert Wyatt, Elizabeth Romero, "Multi-model Ontology-based Hybrid Recommender System in E-Learning Domain," Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 91-95, Sep. 2009.
- [14] Khairil Imran Bin Ghauth, Nor Aniza Abdullah, "Building an E-Learning Recommender System using Vector Space Model and Good Learners Average Rating," Proceedings of the IEEE International Symposium on Advanced Learning Technologies, pp. 194-196, July 2009.
- [15] Robin Burke, "Hybrid Recommender System: Survey and Experiments," User Modeling and User Adapted Interaction, Vol. 12, No. 4, pp. 331-370, 2002.
- [16] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, John Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," Proceeding of the ACM Conference on Computer-Supported Cooperative Work, pp. 175-186, 1994.
- [17] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, "An Introduction to Information Retrieval," Cambridge University Press, pp. 109-133, Apr. 2009
- [18] Inay Ha, Gyu Sik Song, Heung Nam Kim, Geun Sik Jo, "Collaborative Recommendation of Online Video Lectures in e-Learning System," Journal of The Korea Society of Computer and Information, Vol. 14, No. 9, pp. 85-94, Feb. 2009.
- [19] Byung Il Kwon, Nam Mi Moon, "Recommendation System for Supporting Self-directed Learning on E-learning Marketplace," Journal of The Korea Society of Computer and Information, Vol. 15, No. 2, pp. 135-146, Feb. 2010.
- [20] JET, <http://examybmsisa.com>

저 자 소 개



박 재 옥

1999: 서울산업대학교 산업공학과  
공학사  
2006: 서울시립대학교 경영대학원  
이미지니스학과 경영학석사  
2010: 동국대학교 컴퓨터공학과 박사  
수료  
현 재: 새마을금고임직원상조복지회  
영업개발팀 과장  
관심분야: 온톨로지, 데이터베이스,  
e-비즈니스  
Email : ssebbok@dongguk.edu



이 용 규

1986: 동국대학교 전자계산학과 공학사  
1988: 한국과학기술원 전산학과 공학  
석사  
1996: Syracuse University 전산학  
박사  
1978년~83년: 행정직 국가공무원  
1988년~93년: 국방정보체계연구소 선  
임연구원  
1996년~97년: 한국통신 선임연구원  
2002년~03년: 콜로라도대학교 컴퓨터  
학과 방문교수  
1997년~현재: 동국대학교 컴퓨터공학  
과 교수  
관심분야: 데이터베이스, 정보검색,  
웹사이언스, e-비즈니스 시스템  
Email : yklee@dongguk.edu

