

## 이러닝 시스템에서 온라인 비디오 강좌의 협업적 추천 방법

하인애\*, 송규식\*\*, 김흥남\*, 조근식\*\*\*

# Collaborative Recommendation of Online Video Lectures in e-Learning System

Inay Ha\*, Gyu-Sik Song\*\*, Heung-Nam Kim\*, Geun-Sik Jo\*\*\*

### 요약

온라인 비디오 강좌는 내용 파악이 힘든 콘텐츠들이 대부분이기 때문에 학습자가 원하는 정보를 찾기란 쉽지 않다. 그래서 학습자들이 필요로 하는 내용을 정확하고 빠르게 제공해 주는 서비스가 필요하게 되었다. 본 논문에서는 학습자의 요구에 맞는 비디오 강좌를 제공해 주기 위해 사용자 기반의 협업적 여과 방법을 변형하여 적용하고자 한다. 제안하는 알고리즘 방법은 학습자가 평가한 선호도 정보를 바탕으로 강좌의 특성을 이용해 분할한 영역에서 학습자와 비슷한 이웃 학습자들을 찾고, 이웃 학습자들에 의해 높은 선호도를 보인 강좌를 선별하고 강좌의 속성 정보를 반영하여 학습자에게 추천해 주는 방식이다. 즉, 강좌의 특성을 고려하여 강좌별로 분할한 후 사용자 기반의 협업적 여과 방법을 통해 학습자의 선호도를 예측 한다. 그리고 강좌의 속성을 이용한 속성 기반의 여과 방법을 적용해 예측된 강좌들과 유사도를 비교한 후 최종적으로 학습자의 선호도와 가장 유사한 강좌를 추천해 준다.

### Abstract

It is becoming increasingly difficult for learners to find the lectures they are looking for. In turn, the ability to find the particular lecture sought by the learner in an accurate and prompt manner has become an important issue in e-Learning. To deal this issue, in this paper, we present a collaborative approach to provide personalized recommendations of online video lectures. The proposed approach first identifies candidate video lectures that will be of interest to a certain user. Partitioned collaborative filtering is employed as an approach in order to generate neighbor learners and predict learners' preferences for the lectures. Thereafter, Attribute-based filtering is employed to recommend a final list of video lectures that the target user will like the most.

- ▶ Keyword : 사용자 기반 협업적 여과(User based Collaborative Filtering), 속성 기반 여과(Attribute based Filtering), 추천 시스템(Recommendation System), 이러닝(e-Learning), 비디오 강좌(Video Lecture)

• 제1저자 : 하인애

• 투고일 : 2009. 07. 28, 심사일 : 2009. 08. 27, 게재확정일 : 2009. 09. 04.

\* 인하대학교 정보공학과 \*\* 인하대학교 정보컴퓨터교육전공 \*\*\* 인하대학교 정보공학과 교수

## I. 서론

최근 들어 컴퓨터를 이용한 새로운 형태의 교육방법이 제시되면서 이를 위한 이러닝 콘텐츠의 개발이 강조되고 있다. 이미 인터넷을 통해 다양한 교육 프로그램 및 온라인 코스들을 제공하는 사이트가 구축되어 있으며, 이러한 새로운 체제들은 나름대로의 장점을 살려 특성화되고 전문화된 교육을 제공하는 데에 초점을 맞추고 있다. 이에 따라 학습자들은 학교 교육 외에도 자신의 필요에 맞는 교육 프로그램 및 코스 등을 얼마든지 선택하여 학습하는 것이 가능해지고 있다. 온라인 비디오 강좌를 제공하고 있는 대표적인 웹 사이트로, EBS 교육 방송과 메가 스토리, 크레듀, YBM 시사 닷컴, 삼성 SDS 멀티 캠퍼스 등을 볼 수 있다. 이처럼 현재 웹상에는 수많은 종류의 교육 사이트가 존재하고, 이 교육 사이트들은 하루에 수십개씩 새로운 비디오 강좌들을 개설하고 있다. 하지만 온라인 교육 사이트에서 제공하는 비디오 강좌들은 대부분 유료 콘텐츠이므로, 학습자들은 적은 비용으로 자신이 원하는 좋은 강좌를 찾기 위해 많은 시간을 소비하게 된다. 반면에 공급자의 측면에서는 많은 콘텐츠가 생성함에 따라 사용하지 않은 콘텐츠들도 증가하게 됨으로써, 콘텐츠의 낭비를 초래하게 된다. 따라서 공급자가 생성한 콘텐츠들이 적극적으로 활용될 수 있으며, 학습자에게는 필요한 콘텐츠를 정확하고 빠르게 제공해 주는 서비스가 필요하게 되었다. 이러한 요구에 따라 학습자에게 필요한 정보를 제공하는 추천 시스템(Recommender System)이 개발되고 있다. 현재 활성화되어 있는 사이트인 EBS의 경우 상위 10개 랭크된 강좌들을 영역별 또는 과목별로 메인 화면에 위치시켜서 추천해준다. 또한 삼성 SDS 멀티 캠퍼스의 경우 학습자들의 평점을 바탕으로 인기 강좌와 추천 강좌를 별점 형태의 점수와 함께 모든 학습자들에게 제공한다. 이와 같이 온라인 비디오 강좌를 제공하는 사이트들은 개인의 특성에 맞는 추천 방법이 아닌 대다수의 학습자에게 적용되는 추천 방법을 사용한다. 그러므로 특정 학습자가 원하는 강좌를 찾기 위해 소요되는 비용과 시간은 아직도 해결해야 할 문제로 남아있다.

본 논문에서는 이러한 문제점을 해결하기 위해 추천 시스템에서 가장 많이 사용되는 협업적 여과(Collaborative Filtering)와 내용 기반 여과(Content-based Filtering)를 적용하고자 한다. 협업적 여과는 콘텐츠의 과거 구매 기록이나 사용 기록을 이용하여 선별하는 방법이고, 내용 기반 여과는 콘텐츠의 내용을 파악하여 선별하는 방법이다[1]. 본 논문에서는 비디오 강좌를 학습자에게 추천해주기 위해 사용자 기반의 협업적 여과를 변형하여 강좌의 특성 즉, 주제에 따라

분할한 후 M개의 후보 비디오 강좌 리스트를 추출한다. 그리고 비디오 강좌의 속성과 M개의 후보 비디오 강좌를 분석하여 상위 N개의 리스트를 학습자에게 제공해 준다. 이와 같이 학습자의 선호도와 비디오 강좌의 속성을 고려하여 비디오 강좌를 제공하면 수많은 비디오 강좌들 중에서 학습자가 관심 가지는 비디오 강좌를 정확하고 빠르게 제공할 수 있다.

본 논문의 구성은 다음과 같다. 제 2장은 추천 시스템과 기존의 협업적 여과와 내용기반 여과에 대한 배경 지식과 현재 운영되고 있는 사이트에서의 추천방법에 대해서 분석하여 문제점을 소개한다. 제 3장에서는 협업적 여과와 속성기반의 여과를 이용하여 학습자에게 유용한 온라인 비디오 강좌를 추천할 수 있는 시스템의 설계 및 구현에 대해 설명할 것이다. 제 4장에서는 해당 시스템의 실험과 결과를 분석한다. 마지막으로 제 5장에서는 결론과 향후 연구과제에 대하여 정리한다.

## II. 이론적 배경

### 2.1. 추천 시스템

1992년 Goldberg에 의해 제안된 추천 시스템은 사용자들의 관심을 끄는 아이템을 보다 쉽게 찾을 수 있도록 도와주거나 그들의 기호에 기반하여 의미 있는 아이템을 제공해 준다. 이 추천 시스템을 전자 상거래에 접목하여 사용한 기업들 중 대다수가 사용자의 구매를 촉진 시키며 판매가 향상되는 것을 볼 수 있고, 고객 고정 효과도 있음을 알 수 있다. 즉, 고객이 특정 전자상거래 사이트의 추천서비스를 받아들인다면 추천으로부터 얻어지는 이득을 얻기 시작하며 고객의 전환비용은 시간이 지남에 따라 증가하여 다른 경쟁 사이트로의 전환을 꺼리게 될 것이다. 이와 같은 고정 효과는 전자상거래 산업에서 고객의 충성도를 크게 향상시키며 추천시스템의 주요 전략적 가치를 제공하게 된다[2].

추천 시스템은 크게 콘텐츠의 과거 구매 기록이나 사용 기록을 이용하여 선별하는 협업적 여과(Collaborative Filtering)와 콘텐츠의 내용을 파악하여 선별하는 내용 기반 여과(Content-based Filtering)로 나눌 수 있다[1].

#### 2.1.1. 협업적 여과

협업적 여과는 사용자의 관심 사항을 예측하기 위해 유사 사용자의 의견을 사용하여 적합한 정보를 제공하는 방법이다. 즉 아이템에 대한 사용자들의 의견을 데이터베이스에 저장하여 특정 사용자의 특정 아이템에 대한 평가 값을 예측 할 때 특정 사용자와 유사한 사용자들을 발견하고 유사 사용자의 의

견을 참조하여 평가 값을 예측하게 된다[3][4]. 정보자체에 대한 사용자의 선호도 집합을 구성하여 다른 사용자들의 선호도와 비교하게 되며 이를 통해 정보를 여과하게 된다. 즉 동일한 정보에 대한 선호도의 상관계수를 통하여 현 사용자와 기존 사용자들 사이에 기호의 유사성과 그들의 선호도에 기반하여 높은 선호도를 예상하는 정보만 현 사용자에게 여과하여 제공하게 되는 것이다[5]. 예를 들어, 3명의 사용자가 4개의 강좌에 대해 1에서 5점 사이의 점수로 평가(rating)했다고 가정하자. 평가는 특정 사용자가 특정 아이템에 대해서 의견을 표시하는 가장 단순한 방법이다. 표 1.에서 은영이는 선영이보다 은영이와 더 유사하다는 것을 추측할 수 있다. 따라서 은영이가 아직 평가하지 않은 과학에 대한 예측은 선영이보다 은영이를 참조하는 것이 더 낫다고 볼 수 있는데, 이와 같은 방법에 기초한 것이 바로 협업적 여과이다.

표 1. 강좌에 대한 평가 예  
Table 1. Example of learners' evaluation

	국어	영어	수학	과학
은영	5	5	3	2
선영	3	3	5	5
영은	4	3	3	?

협업적 여과는 추천 시스템에서 가장 많이 사용되는 기술로써 Amazon.com, Netflix 등 상업적으로 성공을 거두고 있는 여러 전자상거래 사이트에서 적용하고 있다. 협업적 여과의 장점은 아이템에 대한 사전분류나 파악에 의존하기보다는 다른 사용자들의 의견에 참조한다는 점과 사용자가 예상하지 않았지만 자신의 취향에 맞는 새로운 정보를 발견할 가능성이 있다는 점 그리고 무엇보다도 시스템이 자동화될 가능성이 매우 높다는 점을 들 수 있다[6]. 이러한 방법은 자동화된 프로세스로는 쉽게 분석 될 수 없는 정보의 질을 기존 사용자들의 선호도를 통해 어느 정도 반영한다는 장점이 있으나 정보의 속성에 대한 사용자의 선호도는 고려하지 않기 때문에 속성에 대한 선호도 데이터가 있는 경우에는 이를 활용하지 않는 문제점을 가지고 있다.

### 2.1.2. 내용 기반 여과

내용 기반 여과는 사용자의 프로파일과 추천 아이템의 특징이나 속성(attribute) 사이의 비교를 기반으로 항목을 추천한다. 즉, 아이템 항목과 사용자 선호도 사이의 유사도를 기반으로 하여 추천이 이루어진다. 속성에 대한 사용자의 선호도는 사용자의 명시적인 제시에 의하거나 사용자의 간접적

인 행동을 통해 반영될 수 있다. 이는 사용자의 프로파일을 통해 과거의 구매내역이나 추천 결과를 쉽게 반영한다는 장점이 있다. 이러한 내용 기반 여과는 특정한 단어나 속성 값을 가지는 정보를 여과하기에 매우 효과적이다. 그러나 여과된 정보들의 질을 판별하기에는 적절하지 못하다. 또한 사용자 프로파일에는 없는 속성항목을 가지는 정보에 대한 여과는 수행하기 어렵다[7].

내용 기반 시스템에는 두 가지 추천 방식을 사용할 수 있다. 첫째는 아이템 사이의 특성에 대한 유사성을 이용하는 아이템 상관관계 추천 방식이 있고, 둘째는 사용자 프로파일과 아이템 특성의 유사성을 이용하는 속성기반 추천방식이 존재한다.

그러나 일반적으로 내용기반 추천 시스템은 속성기반 추천 방식을 지칭하는 경우가 많다. 이 시스템은 추천하고자 하는 아이템 자체의 내용과 사용자 프로파일간의 유사성을 고려하여 추천하는 방식이다. 아이템에 관한 정보를 특정 벡터로 표현하고 사용자의 취향을 사용자의 프로파일로 구성한 후, 사용자 프로파일과 아이템 특징 사이의 유사도를 구하여 상품을 추천한다.

내용기반 추천 시스템의 장점은 신문기사나 논문과 같이 내용 정보가 풍부한 아이템에 적용했을 경우 비교적 정확한 추천을 할 수 있다는 것이다. 그러나 비디오 강좌와 같은 멀티미디어 정보처럼 내용 정보가 빈약한 아이템의 경우, 아이템의 특징을 적절히 표현할 수 없기 때문에 유사도 측정 및 올바른 추천이 어렵다는 단점이 있다[8].

## 2.2. 이러닝 추천 시스템

이러닝 추천 시스템은 학습자가 맞춤 교육을 할 수 있도록 제공하는 것으로써, 학습자의 선호도에 따른 학습을 하거나 학습자의 성적이 부진한 과목을 학습할 수 있도록 제시해 준다[9]. 현재 진행 중인 이러닝 추천 시스템의 방법은 콘텐츠 기반의 협업적 여과를 이용한 추천 시스템과 사용자 기반의 협업적 여과를 이용한 추천 시스템, 그리고 하이브리드 여과를 이용한 자동 추천 시스템이 있다[10][11].

콘텐츠 기반의 협업적 여과는 이용한 추천 시스템은 현재 활성화되어 있는 온라인 교육 사이트에서 사용하는 방법으로, 학습자가 과거에 수강한 강좌를 토대로 비슷한 주제의 강좌를 추천해 주는 방법이다. 즉, 목적 학습자가 수강한 과목을 다른 학습자가 동일하게 수강한 경우, 다른 학습자가 수강했던 리스트를 목적 학습자에게 제공해 준다. 이와 같이 학습자가 수강했던 강좌와 유사한 다른 강좌들이 새롭게 업로드 되는 경우 손쉽게 제공받을 수 있지만, 새로운 학습자의 경우 어떠한 강좌도 추천해 줄 수 없는 경우가 생긴다. 이 새로운 학습자가 어떤 강좌에 관심이 있는지 어떠한 정보도 존재하지 않기 때문에 추천이 불가능하다. 이처럼 콘텐츠 기반의 협업적

여과는 사용자의 과거 행동에 종속적이므로 추천받을 수 있는 범위가 한정되어 있다는 단점이 있다.

사용자 기반의 협업적 여과를 이용한 추천 시스템은 먼저 목적 학습자가 강좌를 듣고 평가한 것을 토대로 같은 생각을 가지는 이웃 학습자와의 유사도를 피어슨 상관계수를 이용하여 계산한다. 이때 유사도 값이 큰 이웃 학습자가 수강한 강좌 중 목적 학습자가 수강하지 않은 리스트를 추출하여 목적 학습자에게 추천해 준다(10). 이와 같이 유사한 이웃 학습자들로부터 추천을 받는 경우, 불필요한 강좌의 추천이 감소되어 학습자에게 신뢰도를 향상시킬 수 있다. 하지만 학습자의 학습 활동이 활발하지 않은 경우, 이웃 학습자의 구성이 쉽지 않아 바른 추천이 어렵다는 단점이 있다. 그래서 학습자뿐 아니라 이템까지도 고려하는 하이브리드 추천 시스템이 연구되어지고 있다.

하이브리드 추천 시스템은 학습자의 로그를 이용하여 학습자 프로파일을 생성하고 협업적 여과를 통하여 관심사가 유사한 이웃 학습자들을 파악하게 된다. 그리고 강좌를 텍스트 마이닝하여 생성된 콘텐츠 프로파일을 콘텐츠 기반의 여과 방법을 적용하여 새로운 리스트를 추천한다. 이렇게 2가지의 여과 방법을 통해 추천된 리스트를 혼합하여 상위 N개의 리스트를 최종적으로 학습자에게 추천해준다(11). 이와 같이 학습자와 강좌 둘 다 고려하여 학습자에게 추천해줌으로써 신뢰성을 회복시킬 순 있지만, 고관여 서비스(High-involvement Service)인 이러닝에는 적합하지 못하다. 즉, 이러닝은 충분한 시간을 두고 의사 결정을 해야 하는 것으로 충분한 정보 탐색과 신중한 평가를 필요로 한다.

### III. 협업적 온라인 비디오 강좌 추천 시스템

#### 3.1. 시스템 구조

본 논문에서는 그림 1.에서 보는바와 같이 학습자의 선택에 영향을 주는 강좌의 특성(주제, 과목, 분야)을 고려하여 연관성이 있는 강좌들로 구성된 분할된 집합을 구성하여 일괄적인 협업적 여과가 아닌 분할된 협업적 여과를 진행한다. 그리고 정보의 속성에 대한 학습자의 선호도를 고려하지 못하는 기존 협업적 여과 알고리즘의 문제점을 개선하고자 정보의 속성에 대한 학습자의 선호도를 분할된 협업적 여과에 반영함으로써 예측의 정확성을 향상시키는 방법을 제안한다.

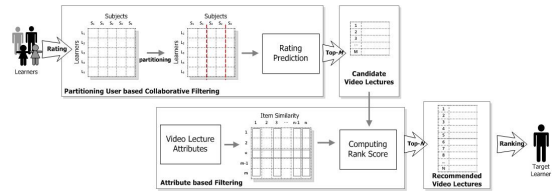


그림 1. 시스템 흐름도  
Fig. 1. System flow

시스템은 크게 분할된 협업적 여과와 속성 기반 여과로 나누어진다. 첫 번째 과정인 분할된 협업적 여과는 다른 학습자들의 선호도 정보를 바탕으로 유사한 성향을 가지는 이웃 학습자들을 찾고, 그 이웃 학습자들에 의해 높은 선호도를 보인 추천 후보 강좌를 선별한다. 그 방법으로는 먼저 최초의 입력 값으로써 학습자-강좌 선호도 행렬을 구성한다. 그리고 분할된 협업적 여과 방법으로 입력된 학습자-강좌 선호도 행렬을 각 주제에 따라 각각의 행렬로 분할화 한다. 각각의 분할된 학습자-강좌 선호도 행렬을 대상으로 피어슨 상관계수를 이용하여 학습자의 이웃을 구성한다. 마지막으로 학습자 기반의 선호도 예측을 하여 M개의 후보 강좌 리스트를 결과 값으로 출력한다.

두 번째 아이템 속성 기반의 정보를 이용하여 강좌간의 유사도를 계산하고 분할된 협업적 여과를 통해 선정된 M개의 후보 강좌들 중에서 최종적으로 학습자에게 추천할 N개의 강좌들을 추출한다. 일반적으로 후보 강좌의 수 M은 추천할 강좌의 수 N보다 작다.

#### 3.2. 분할된 협업적 여과 방법

일반적으로 온라인 강좌 서비스를 제공하는 사이트는 학습자들이 흥미를 가질 수 있는 강의를 제공하기 위해, 한 강좌의 수강이 끝나면 강의가 어땠는지 평가할 수 있도록 되어 있다. 이와 같이 이미 수강한 학습자들의 평가 정보는 새로운 학습자들의 온라인 강좌 선택에 대한 의사 결정에 영향을 미친다. 온라인 강좌 서비스를 제공하는 사이트를 통해 학습자들의 강좌 수강 패턴을 분석한 결과 강좌의 주제에 따라 선호도의 평가가 달라지는 경향을 보였다. 따라서 본 논문에서는 학습자-강좌 행렬에서 보다 세부적으로 주제에 따라 분할하여 그림 2.와 같이 주제에 따른 행렬을 생성한다. 즉, 학습자가 비디오 강의를 시청한 후 해당하는 비디오 강좌에 대해서 평가한 것 중에서 성격이 비슷한 비디오 강좌를 묶어 재구성 한다.

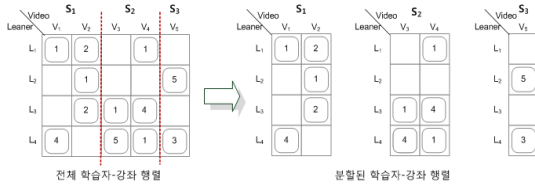


그림 2. 주제별 행렬 분할화

Fig. 2. Matrix partitioning according to subjects

### 3.2.1. 분할된 행렬에서 학습자 이웃 구성

협업적 여과 방법에서 가장 중요한 단계로써 특정 사용자와 유사한 선호도를 가지는 이웃들의 집합을 찾는 과정이다. 유사도의 계산은 통계적인 방법 중 일반적으로 많이 사용되어지는 피어슨 상관계수(Pearson correlation coefficient)를 분할된 행렬에 적용하여서 다음 식(1)과 같이 사용한다 [12][13].

$$Sim_s(u, v) = \frac{\sum_{j \in CRI(s)} (r_{u,j} - avg_s(r_u)) (r_{v,j} - avg_s(r_v))}{\sqrt{\sum_{j \in CRI(s)} (r_{u,j} - avg_s(r_u))^2 \sum_{j \in CRI(s)} (r_{v,j} - avg_s(r_v))^2}} \quad \text{식(1)}$$

여기서  $sim_s(u, v)$ 는 학습자  $u$ 와  $v$ 가 분할된 행렬( $s$ )에서 공통으로 평가한 강좌의 선호도를 바탕으로 계산한 두 학습자 간의 유사도를 나타낸다.  $CRI(s)$ 는 주제  $s$ 의 아이템에 대해서 학습자  $u, v$ 가 동시에 평가한 아이템의 집합을 의미한다. 그리고  $r_{u,j}$ 와  $r_{v,j}$ 는 두 학습자  $u$ 와  $v$ 의 강좌  $j$ 에 대한 선호도 값이고,  $avg_s(r_u)$ 와  $avg_s(r_v)$ 는 사용자  $u$ 와  $v$ 가 분할된 행렬( $s$ )에서 공통으로 평가한 모든 강좌의 선호도 평균값이다.

일반적인 협업적 여과 방법에서 사용하는 피어슨 상관계수는 전체 사용자와 아이템의 행렬에 대해서 계산하지만 본 논문에서는 아이템(비디오 강좌)을 주제별로 구분한 분할된 사용자 아이템 행렬에서 계산하여, 분할한 행렬의 특성을 반영할 수 있도록 한다.

### 3.2.2. 학습자 기반 강좌 선호도 예측

학습자 기반 강좌 선호도 예측은 각각의 학습자들이 학습하지 않은 강좌에 대해 선호도 평가 값을 예측하는 과정이다. 이때 이웃에 속한 학습자들의 선호도 평가 값의 가중치 합을 통해서 예측이 이루어진다. 선호도 예측 값은 다음 식(2)와 같이 계산한다.

$$P_{u,j} = avg_s(r_u) + \frac{\sum_{v \in KNN_s(u)} sim_s(u, v) (r_{v,j} - avg_s(r_v))}{\sum_{v \in KNN_s(u)} sim_s(u, v)} \quad \text{, if item } j \in \text{subject } s \quad \text{식(2)}$$

이 때,  $avg_s(r_u)$ 와  $avg_s(r_v)$ 는 각 학습자  $u$ 와  $v$ 가 분할된 행렬( $s$ )에서 평가한 선호도 평균을 의미하고,  $r_{u,j}$ 는 사용자  $v$ 가 강좌  $j$ 에 대해 갖는 선호도 평가 값을 나타낸다. 그리고

$Sim_s(u, v)$ 는 분할된 행렬( $s$ )에서의 유사도 값을 의미하며,  $KNN_s(u)$ 는 주제  $s$ 에 대한 학습자  $u$ 의 유사한 이웃 학습자 집합을 의미한다.

다음 알고리즘 1.은 위의 분할된 행렬에서 이웃을 구성하고 선호도를 예측해  $M$ 개의 후보 비디오 강좌 리스트를 추출해 내는 알고리즘이다.

알고리즘 1. M개의 후보 비디오 강좌 리스트 식별

Algorithm 1. Extraction of candidate video lectures

```

Input: target user, u; items not rated by user u, Lu; a number of candidated items, M; a number of neighbors, k;
Output: a set of candidated items for user u CMu

01 candidatedItemsGeneration(u, Lu, M, k)
02 for each j ∈ Lu
03   s ← subject of item j
04   for v ← 1 to n
05     CRI(s) ← a set of items in subject s rated
06     by both user u and v
07     sims(u, v) ←
08     partitionedPearsonCorrelation(u, v, CRI(s))
09     Sort sims(u, v)
10     KNNs(u) ← k nearest neighbors to user u
11     according to
12     sims(u, v)
13     for each v ∈ KNNs(u)
14       avgs(rv) ← the average rating of user v in
15       subject s
16       p1 ← p1 + sims(u, v) × (rv,j - avgs(rv))
17       p2 ← p2 + | sims(u, v) |
18       avgs(ru) ← the average rating of user u in
19       subject s
20       pu,j ← avgs(ru) + p1/p2
21   for each j ∈ Lu
22     if pu,j = among the M highest values in pu,*
23   then
24     CMu ← CMu ∪ { j }
25 return CMu
    
```

### 3.3. 속성 기반 여과 방법

학습자에게 필요한 온라인 비디오 강좌를 추천해 주기 위해 표 2.와 같이 강좌의 다양한 속성 정보를 참조한다. 즉, 온라인 강좌를 강의하는 선생님이 누구이며, 강좌의 주제 또는 과목은 무엇이며, 강좌의 난이도와 구성은 어떻게 되어있는지 등 강좌의 특성을 알 수 있는 다양한 정보를 온라인 비디오

강좌는 가지고 있다.

표 2. 온라인 비디오 강좌의 속성 예  
Table 2. Attribute example of online video lecture

강좌	선생님	과목	난이도	이론구성	실습구성
V1	Jennifer	국어	중상	4	6
V2	Inay	국어	중	3	7
V3	Daniel	수학	하	7	3
V4	Nami	수학	상	2	8
V5	James	과학	중상	4	6

하지만 일반적인 협업적 여과 방법을 이용하면 학습자의 강좌에 대한 선호도 평가 값만을 이용하게 되기 때문에 학습자 입장에서 중요한 강좌의 특성에 대한 선호도를 파악하기 어렵고 이런 중요한 정보를 반영하여 추천하기 어려운 문제점이 있다. 따라서 본 논문에서는 이러한 강좌의 속성 정보를 이용하여 속성 기반의 여과 방법을 통해 분할된 협업적 여과와 혼합하여 각각의 학습자에게 추천하고자 한다. 속성 기반의 여과 방법 순서는 먼저 속성 기반 아이템 유사도 측정을 진행하고, 유사도에 기반을 두어 선호도를 예측하여 학습자에게 추천 한다.

속성 값을 이용하여 아이템 간의 유사도를 측정하기 위해 각 속성의 특성에 맞게 속성 값을 수치화 하였다. 강좌의 속성 중 선생님과 과목은 스트링 비교를 통하여 같으면 1, 다르면 0이 된다. 난이도 속성은 공통, 중급, 중상, 고급 등으로 나누어지며, 1.0, 2.0, 2.5, 3.0으로 수치화 하였다. 그 외에 다른 속성들도 위와 비슷한 방법으로 속성 값을 수치화 하여 속성 기반 여과 방법을 진행하였다.

### 3.3.1. 속성 기반 강좌 랭킹

분할된 협업적 여과로부터 선별된 후보 비디오 강좌들은 강좌의 속성을 고려하여 랭킹 점수를 구한다. 랭킹 점수는 특정 강좌의 속성과 유사한 강좌에 대해서 특정 학습자가 과거에 얼마나 선호한 강좌였는지를 참조하여 계산한다. 학습자  $u$ 의 강좌  $j$ 에 대한 랭킹 점수  $Score_{u,j}$ 는 다음 식(3)과 같이 정의된다.

$$Score_{u,j} = \sum_{i \in KMS(j)} attr(i,j) \times r_{u,i}, \quad j \leftarrow CM_u \quad \text{식(3)}$$

$KMS(j)$ 는 강좌  $j$ 와 가장 유사한 아이тем 집합을 의미하며,  $r_{u,i}$ 는 학습자  $u$ 가 강좌  $i$ 에 대한 평가 값 정보를 나타낸다. 그리고  $attr(i,j)$ 는 강좌  $i$ 와 후보 강좌 집합  $CM_u$ 에 있는  $j$ 와 의 속성 기반의 유사도를 말하며, 이는 두 강좌의 속성에 대

해서 코사인 유사도로 계산한다.

알고리즘 2.는 속성 기반 여과를 이용하여 랭킹 점수가 높은  $N$ 개의 강좌를 추천하는 과 강좌보여준다.

알고리즘 2. 상위  $N$ 개의 비디오 강좌 추천  
Algorithm 2. Recommendation of video lectures

```

Input: target user,  $u$ ; total item list,  $I$ ; candidated items for user  $u$ ,  $CM_u$ ; a number of recommended items,  $N$ ; a number of most similar items,  $k'$ 
Output: a set of recommended items for user  $u$ 
TopNu

01 recommendingItems( $u, I, CM_u, N, k'$ )
02 for each  $j \in CM_u$ 
03   for each  $i \in I$ 
04      $attr_i(i, j) \leftarrow \text{attributeSimilarity}(i, j)$ 
05   Sort  $attr_i(i, j)$ 
06    $KMS(j) \leftarrow k'$  most similar items to item  $j$ 
07   according to  $attr_i(i, j)$ 
08   for each  $i \in KMS(j)$ 
09      $score_{u,j} \leftarrow score_{u,j} + attr_i(i, j) \times r_{u,i}$ 
10
11   for each  $j \in CM_u$ 
12     if  $score_{u,j}$  is among the  $N$  highest ranking
13     score in  $score_{u,*}$  then
14        $TopN_u \leftarrow TopN_u \cup \{j\}$ 
15   return  $TopN_u$ 
    
```

## IV. 실험 및 평가

### 4.1. 실험 평가 방법

본 논문에서 제안한 사용자 기반 분할된 협업적 여과와 속성 기반 여과를 이용한 방법(OVRS)의 성능을 비교하기 위해 전형적인 사용자 기반 협업적 여과 방법(UBCF)과 분할된 협업적 여과 방법(PUBCF)을 이용하였다. 실험 데이터는 2008년도 EBS 교육방송 중 내신영역에 속하는 국어, 수학, 영어, 사회, 과학의 5과목에 대한 61개의 비디오 강좌를 사용하였다. 그리고 본 실험에 고등학교 3학년 학생 349명이 참여하여 각 강좌에 대해서 최소 1점 ~ 최고 5점으로 평가하였고, 학습자들이 평가한 정보는 2,416개이다.

선호도 예측 정확성 평가를 위해 사용자들이 이미 평가한 아이тем들에 대하여 "All but 1" 프로토콜을 사용하여 예측하

였다. "All but 1"은 사용자가 이미 평가한 아이템 중 임의로 하나의 아이템을 선택하여 테스트로 사용하고, 이것을 제외한 나머지 데이터를 이용하여 테스트 아이템에 대한 예측을 수행하는 방법이다[1]. 이 선호도 예측 값의 정확성을 평가하기 위해 식(4)와 같이 평균 절대 에러(MAE : Mean Absolute Error)를 사용하였다.

$$MAE = \frac{\sum_{i=1}^N |r_i - p_i|}{N} \quad \text{식(4)}$$

위의 식에서  $N$ 은 총 예측 횟수이며  $r_i$ 는 실제 값, 그리고  $p_i$ 는 예측 값을 나타낸다. 평균 절대 에러는 각 예측 단계  $i$ 에서의 실제 값과 예측 값의 차로 계산한 오차의 합을 총 예측 횟수로 나누어 계산한다. 이미 평가된 값이 존재하지만 점수를 평가하지 않은 것으로 가정하고 그 값을 가린 다음 다른 데이터를 이용하여 점수를 예측한다.

다음으로 의사 지원 정확성 평가를 위해 표 3.과 같이 추천된 강좌와 추천되지 않은 강좌의 데이터를 이용해 평가하였다[14].

표 3. 의사 지원 정확성 평가 데이터  
Table 3. Evaluation of decision support accuracy

	추천된 강좌	추천되지 않은 강좌	선호도 범위
relevant	Nr	Nnr	4,5
irrelevant	Ni	Nni	1,2

평가 방법으로는 추천되어진 강좌  $N$ 개가 사용자들에게 얼마나 정확하게 추천되어졌는지 알아보기 위해 식(5)와 같이 정확율(Precision)과 재현율(Recall)을 사용하였다. 정확율은 추천된 강좌들 중에서 선호도가 4 또는 5의 값을 가지는 강좌의 수를 나타내며, 재현율은 높은 선호도 값을 가지는 모든 강좌들 중에서 추천된 강좌의 수를 나타낸다

$$Precision = \frac{N_r}{N_r + N_i}, \quad Recall = \frac{N_r}{N_r + N_{nr}} \quad \text{식(5)}$$

위 식에서  $N_r$ 은 추천된 강좌 중 높은 선호도인 4, 5의 값을 가지는 강좌 수이고,  $N_i$ 는 추천된 강좌 중 낮은 선호도인 1, 2의 값을 가지는 강좌 즉 잘못 추천된 강좌들의 수를 의미한다. 그리고  $N_{nr}$ 은 선호도 값이 4, 5인 강좌들 중 추천되지 않은 강좌의 수를 나타낸다.

정확율과 재현율은 사용자에게 적절한 아이템들이 추천되었는지 측정하는 평가 방법으로 이들의 값이 크면 클수록 추

천 성능이 우수함을 의미한다. 그러나 정확율과 재현율은 서로 반비례 관계가 존재한다. 예를 들어  $N$ 이 증가하면 재현율은 증가하고 정확율은 감소하는 경향이 있다. 그러므로 정확율과 재현율에 동일한 가중치를 주어 하나로 통합한 F1-measure 방법도 사용한다. F1-measure는 다음 식(6)과 같다.

$$F1\text{-measure} = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad \text{식(6)}$$

## 4.2. 실험 결과 및 분석

### 4.2.1. 분할된 협업적 여과 성능 효과

본 논문에서 제안한 분할된 협업적 여과 성능의 효과를 평균 절대 에러 실험을 통해 알아보았다. 평균 절대 에러 실험은 표 4.와 같이 목적 사용자의 유사한 이웃 사용자의 크기  $k$ 를 20, 40, 60, 80, 100으로 증가하여 사용자 기반의 협업적 여과(UBCF)와 분할된 사용자 기반의 협업적 여과(PUBCF)를 비교하였다.

표 4. 평균 절대 에러 실험 결과  
Table 4. Experimental result for MAE

k	20	40	60	80	100
UBCF	0.3806	0.3722	0.3730	0.3783	0.3802
PUBCF	0.2929	0.2884	0.2942	0.2999	0.2999

표 4.의 실험결과를 보면 유사한 이웃 사용자의 수가 40일 때 UBCF와 PUBCF가 각각 0.3722와 0.2884로 가장 좋은 성능을 보이는 것을 볼 수 있다. 하지만, 유사한 이웃 사용자 수의 변화에 따라 큰 성능의 차이를 보이지 않는 것을 볼 수 있다. 또한 전체적으로 사용자 기반의 협업적 여과 보다 분할된 협업적 여과 방법을 사용하는 것이 0.7% ~ 0.8% 정도 향상된 선호도 예측을 할 수 있음을 볼 수 있다. 그러므로 강좌의 특성을 고려하여 행렬을 분할한 후 협업적 여과 방법을 사용했을 때, 모든 강좌에 대한 협업적 여과 방법을 사용하는 것 보다 향상된 성능을 가져올 수 있다는 결과를 확인할 수 있다.

### 4.2.2. 속성 기반 여과 성능 효과

강좌의 속성들을 이용하여 여과하는 경우 얼마나 정확하게 예측을 할 수 있는지 실험을 통하여 비교하였다. 실험은 분할된 협업적 여과 성능 실험을 통해 좋은 결과를 가져온 40과 60일 때의 유사한 사용자의 수를 이용하였고, 후보 강좌 리스트의 수  $M$ 을 15로 고정하였다. 그리고 유사한 강좌의 수  $k'$

를 10, 15, 20으로 증가하여 비교하였다.

표 5. 유사한 강좌의 수에 따른 추천 결과  
Table 5. Recommendation result according to a number of similar lectures  
(Neighbor=40, M=15)

k'	10		15		20	
	추천 강좌	추천 안된 강좌	추천 강좌	추천 안된 강좌	추천 강좌	추천 안된 강좌
relevant	521	559	520	560	496	584
irrelevant	8	438	9	437	9	437

표 5.는 유사한 이웃 사용자의 수 k가 40일 경우에 사용자가 선호하는 강좌를 추천해 주는지 나타내는 표이다. 위의 표에서 유사한 강좌의 수 k가 10인 경우에 사용자의 선호도와 관련이 있는 강좌를 추천해 주는 것을 볼 수 있다. 그리고 사용자의 선호도와 관련이 적은 강좌의 추천도 k가 10인 경우 적게 추천되는 것을 볼 수 있다. 즉, k가 15와 20일 때 사용자에게 잘못 추천해 주는 경우가 더 많이 발생하는 것을 알 수 있다.

표 6. 유사한 강좌의 수에 따른 추천 결과  
Table 6. Recommendation result according to a number of similar lectures  
(Neighbor=60, M=15)

k'	10		15		20	
	추천 강좌	추천 안된 강좌	추천 강좌	추천 안된 강좌	추천 강좌	추천 안된 강좌
relevant	531	549	538	542	519	561
irrelevant	7	439	8	438	7	439

표 6.은 유사한 이웃 사용자의 수 k가 60일 때 얼마나 사용자의 선호도에 맞는 강좌를 추천해 주는지 비교한 표이다. 위의 경우 유사한 강좌의 수 k가 15일 때 사용자가 높이 선호하는 강좌들을 추천해 주는 것을 볼 수 있다. 반면에 k가 15인 경우 사용자의 선호도와 관련이 없는 강좌들도 많이 추천되어지는 것을 볼 수 있다.

그러므로 표 5.와 표 6.의 결과를 봤을 때, 사용자의 선호도에 맞는 강좌를 정확하게 추천해 주기 위해서는 유사한 강좌의 수를 10으로 하는 경우 좋은 결과를 가져올 수 있음을 알 수 있다.

### 4.2.3. 추천 성능 비교 평가

본 논문에서 제안한 여과 방법의 추천 성능 평가를 위해 각각의 알고리즘에 대하여 재현율, 정확율, F1-measure를 비교하였다. 실험은 앞의 분할된 사용자 기반의 협업적 여과 방법과 속성 기반 여과 방법의 실험을 통해 얻은 결과 값을 기본으로 이용하였다. 즉, 유사한 이웃 사용자의 수 k는 40으로 하고, 유사한 강좌의 수 k'는 10으로 고정시켰다. 그러므로 UBCF와 PUBCF는 유사한 이웃 사용자의 수를 40으로 하였고, 본 논문에서 제안한 PUBCF+ABF는 유사한 이웃 사용자의 수 40과 유사한 강좌의 수 10을 혼합하여 실험하였다.

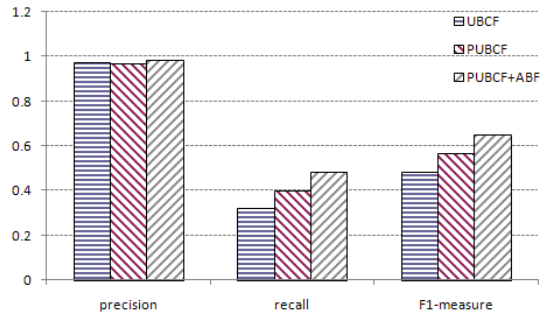


그림 3. 알고리즘별 추천 성능 비교  
Fig 3. Comparison of recommendation performance

그림 3.은 3개의 알고리즘별로 정확율과 재현율 F1-measure의 성능을 비교한 결과이다. 먼저 정확율을 보면 미세한 차이지만 본 논문에서 제안한 PUBCF+ABF가 다른 두 알고리즘에 비해 높은 정확율을 보이는 것을 볼 수 있다. 본 실험을 통해 학생이 문과에 속하는지 이과에 속하는지에 따라 사용자에게 추천해 주는 과목이 현저히 다른 것을 볼 수 있었다. 즉, 문과 학생의 경우 사회 과목의 강좌를 선호하지만 이과 학생의 경우에는 과학 과목을 선호하는 경향이 나타난다. 그리고 대부분 문과 학생들은 과학 과목을 한 과목도 수강하지 않았고, 반대로 이과 학생들은 사회 과목을 한 과목도 수강하지 않은 것을 알 수 있었다. 이런 경우 실험 결과 문과 학생들에게 과학 과목은 모든 강좌가 0점으로 예측되고, 이과 학생들에게는 사회 과목이 0점으로 예측된다. 그래서 문과 학생들에게 과학 과목을 추천해 주거나 이과 학생들에게 사회 과목을 추천해야 하는 경우 분할된 사용자 기반의 협업적 여과 방법보다 전 과목을 대상으로 하는 일반적인 사용자 기반의 협업적 여과 방법을 이용해야 보다 좋은 성능을 가져올 수 있다. 재현율협업적 여과도 예측에 이용할 수 있는 실험 데이터들이 희박하였기 때문에 전반적으로 재현율 값이 낮은 것을 볼 수 있다. 재현율도 정확율의 결과 값과



마찬가지로 본 논문에서 제안한 PUBCF+ABF의 알고리즘이 48%로 가장 높은 값을 나타낸다. 또한 일반적인 사용자 기반의 협업적 여과보다 분할된 사용자 기반의 협업적 여과가 약 8% 정도 더 좋은 재현율을 나타내는 것을 확인할 수 있다. 마지막으로 정확율과 재현율에 동일한 가중치를 주어 하나로 통합한 F1-measure의 성능을 보면, 정확율과 재현율의 결과와 마찬가지로 본 논문에서 제안한 PUBCF+ABF가 64%로 가장 좋은 성능을 보였다. 또한 일반적인 사용자 기반의 협업적 여과보다 분할된 사용자 기반의 협업적 여과가 8% 정도 성능이 향상되는 것을 볼 수 있다.

종합적으로 실험 결과를 분석해보면, 모든 평가 방법에서 분할된 사용자 기반의 협업적 여과 방법과 속성 기반 여과 방법을 혼합한 경우 가장 성능이 좋았고, 분할된 사용자 기반의 협업적 여과 방법이 그 다음으로 성능이 좋음을 볼 수 있었다. 즉, 일반적으로 많이 사용되어지는 사용자 기반의 협업적 여과 방법보다 본 논문에서 제안한 분할된 사용자 기반의 협업적 여과 방법과 속성 기반 여과 방법을 혼합하여 사용하는 경우 사용자에게 바른 추천을 할 수 있음을 알 수 있다.

## V. 결론 및 향후 연구

웹의 발달로 인해 온라인 교육이 활성화되어 있는 요즘에는 수많은 온라인 비디오 강좌로 인해 사용자들이 자신의 요구에 맞는 강좌를 선택하는 것이 쉽지 않다. 그러므로 본 논문에서는 내용 파악이 쉽지 않은 온라인 비디오 강좌의 효과적인 추천을 위해 분할된 협업적 여과와 속성 기반 여과를 혼합한 여과 방법을 제안하였고, 이 여과 방법이 온라인 비디오 강좌 추천 시스템에 효율적임을 알 수 있었다. 본 논문에서 제안한 여과 방법은 사용자-강좌 행렬에서 강좌의 특성이나 주제에 따라 분할된 사용자 기반의 협업적 여과 방법과 강좌의 속성을 분석하여 적용한 속성 기반의 여과 방법을 접목시킨 것이다. 이 여과 방법은 일반적인 사용자 기반의 협업적 여과 방법보다 세부적인 것으로, 전체보다는 일부분에 초점을 맞추어 사용한 기법이다. 제안한 여과 방법의 성능을 평가하기 위해 일반적으로 널리 알려진 상위  $N$ 개의 추천 방법을 이용하였으며, 여기서 얼마나 정확하게 추천이 되어졌는지 정확율과 재현율 그리고 F1-measure를 통해 살펴보았다. 그 결과 일반적으로 많이 사용되어지는 사용자 기반의 협업적 여과 방법을 사용하는 것보다 강좌의 특성을 반영하여 분할된 사용자 기반의 협업적 여과 방법을 사용하는 것이 추천 성능에 효율적임을 볼 수 있었다. 또한 분할된 사용자 기반의 협업적 여과 방법에 강좌의 속성들을 반영한 속성 기반 여과 방법을

혼합하여 사용하는 경우 더 정확하게 추천되어지는 것을 볼 수 있다. 하지만 성능을 평가하는데 사용된 데이터의 수가 너무 미약하였으므로, 향후 연구에는 보다 다양한 데이터를 여러 방법을 통해 수집할 필요성이 있다. 그리고 보다 좋은 재현율을 위해 다양한 강좌 속성도 고려해야 한다. 또한 본 논문의 실험에 참여한 사용자가 349명으로 소규모에 속하였으므로, 실제 온라인에서 학습하는 사용자들의 수처럼 더 많은 사용자들이 참여할 수 있는 환경이 필요하다.

## 참고문헌

- [1] Breese, J. S., Heckerman, D. and Kadie, C. "Empirical Analysis of Predictive Algorithms for Collaborative Filtering." In Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence, pp. 43 -52. May. 1998.
- [2] Basu, C., Hirsh, H. and Cohen, W. W. "Recommendation as classification: using social and content-based information in recommendation." In Proceedings of the 1998 National Conference on Artificial Intelligence (AAAI-98), pp. 714-720. 1998.
- [3] Tang, Y. T. and McCalla, G. "Mining the implicit ratings for focused collaborative filtering for paper recommendations." In Proceedings of 9th International Conference on User Modeling, 2003.
- [4] Bell, R. M. and Koren, Y. "Improved Neighborhood-based Collaborative Filtering" In Proceedings of KDD'07, Aug. 2007.
- [5] Pennock, D. M., Horvitz, E., Lawrence, S. and Giles, L. "Collaborative Filtering by Personality Diagnosis: A Hybrid Memory - and Model-Based Approach." In Proceeding of the 6th Conference on Uncertainty in Artificial Intelligence, pp. 473-480. 2000.
- [6] Bobadilla, J., Serradilla, F., Hermando, A. "Collaborative filtering adapted to recommender systems of e-learning." Knowledge-Based System, Vol. 22, Issue. 4, pp. 261-265. May. 2009.
- [7] Basilico, J. and Hofmann, T. "Unifying Collaborative and Content-Based Filtering." In Proceedings of the 21st International

Conference on Machine Learning, 2004.

[8] Funakoshi, K. and Ohguro, T. "A content-based collaborative recommender system with detailed use of evaluations," In Proceedings of 4th International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, Vol. 1, pp. 253-256. 2000.

[9] 김천식, 정명희, "IRT와 데이터 마이닝을 이용한 효과적인 평가 및 추천시스템". 한국컴퓨터정보학회 논문지, 제11권, 4호, 109-117쪽, 2006년 9월

[10] Tan, H. and Guo, J. "E-Learning Recommendation System," In Proceedings of the 2008 International Conference on Computer Science and Software Engineering, vol. 5, pp. 430-433. Dec. 2008.

[11] Khribi, M. K., Jemni, M., Nasraoui, O. "Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval," In Proceedings 8th IEEE International Conference on Advanced Learning Technologies, pp. 241-245. Jul. 2008.

[12] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. "GroupLens: An open architecture for collaborative filtering of netnews," Proceedings of ACM CSCW'94 Conference on Computer Supported Cooperative Work, pp. 175-186. Mar. 1994.

[13] Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R. and Riedl, J. "GroupLens: Applying collaborative filtering to Usenet news," Communications of the ACM, Vol. 40, No. 3, pp. 63-65. Mar. 1997.

[14] Herlocker, J., Konstan, J., Terveen, L. and Riedl, J. "Evaluating Collaborative Filtering Recommender Systems," ACM Transactions on Information Systems, Vol. 22, pp. 5-53. Jan. 2004.

## 저자 소개



### 하 인 애

2005 수원대학교 공학사.  
2007 인하대학교 공학석사.  
2007 - 현재 인하대학교 박사과정  
관심분야 : 추천시스템, 이러닝, 시맨틱 웹, 개인화 시스템 등



### 송 규 식

2006 인하대학교 공학사.  
2009 인하대학교 공학석사.  
2008-현재 (주) 비퍼엔알 SAP  
Netweaver Consultant  
관심분야 : 추천시스템, 이러닝, 시맨틱 웹 등



### 김 흥 남

2002 인하대학교 공학사  
2004 인하대학교 공학석사  
2009 인하대학교 공학박사  
2009-현재 Postdoctoral research,  
University of Ottawa  
관심분야 : 추천시스템, 데이터 마이닝, 시맨틱 웹 등



### 조 근 식

1982 인하대학교 공학사  
1985 Queens College/City  
University of New York  
(MA in Computer Science)  
1991 City University of New  
York (Ph. D in Computer  
Science)  
1991-현재 인하대학교 컴퓨터정보공  
학과 교수  
2006-현재  
인하대학교 BK21 지능형  
유비쿼터스 물류기술 연구사업단장  
관심분야 : 인공지능, 온톨로지,  
CSP, 전자상거래 등