



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

교 육 학 석 사 학 위 논 문

**협업 필터링 기법을 이용한 추천  
시스템**



이 논문을 교육학석사 학위논문으로 제출함.

2012년 11월

부 경 대 학 교 교 육 대 학 원

전 산 교 육 전 공

김 은 숙



## [차 례]

차례.....	i
표 차례 .....	ii
그림, 수식 차례 .....	iii
Abstract .....	iv
I. 서 론 .....	1
II. 관련 연구 .....	3
2.1 추천 시스템.....	3
2.2 추천 기술의 종류 .....	5
2.3 협업 필터링 기법 .....	7
2.4 협업 필터링 기법의 문제점 .....	16
III. 설계 및 구현 .....	17
3.1 카테고리 세분화 협업 필터링 추천 시스템 구성 .....	17
3.2 카테고리 세분화 협업 필터링 단계 .....	19

IV. 실험 및 성능 평가 .....	21
4.1 카테고리 세분화 협업 필터링 추천 실험 .....	22
4.2 실험 평가 .....	28
V. 결론 및 향후 연구 .....	30
참고문헌 .....	32

**[표 차례]**

표 1. 추천기술의 종류 .....	6
표 2. 협업 필터링 추천의 예 .....	9
표 3. 컨텐츠 대분류/소분류 구분 정보 .....	22
표 4. 목표고객 평가치 .....	23
표 5. 사용자 20명 평가치 .....	23
표 6. 목표고객과 사용자간 유사도 .....	24
표 7. 대분류 카테고리간 유사도 .....	25
표 8. 대분류 선호도 예측 값 .....	25

표 9. 소분류 카테고리간 유사도 .....27

표 10. 소분류 선호도 예측 값 .....28



## [그림 차례]

그림 1. 추천 시스템 .....	4
그림 2. 협업 필터링 추천 기법 .....	8
그림 3. 협업 필터링 추천 과정 .....	12
그림 4. 추천 시스템 구조 .....	18
그림 5. 목표고객 소분류 카테고리 선택 정보 .....	26
그림 6. 사용자 소분류 카테고리 선택 정보 .....	26
그림 7. 날짜 경과에 따른 목표고객 선택 횟수 .....	29

## [수식 차례]

수식 1. 피어슨 상관계수 공식 .....	13
-------------------------	----

수식 2. 코사인 유사도 공식 ..... 13

수식 3. 선호도 예측 공식 ..... 14

수식 4. 아이템 기반 선호도 예측 공식 ..... 15



# Using Collaborative Filtering Techniques Recommendation System

Eun Sook Kim

*Graduate School of Education*

*Pukyong National University*



**ABSTRACT**

Since advertising market has recently grown rapidly, providing items quickly and accurately to customers is a factor of success. A recommendation system helps to provide content to meet the needs of these individuals. It often takes a long time for each individual to make a decision when they want to choose a product because of the wide range of content provided. To solve this problems,

the content that users want should be predicted. A system that allows users to select content within a short time is important.

This study shows how a selection system can increase the number of referrals using collaborative filtering techniques in order to provide content to meet users' needs. In the first step, the selection is divided into main categories and then similarities are sought between targeted customers and users. The most relevant main category was recommended to the customer by first obtaining the predicted affinity for the main categories and by configuring the nearest neighbors. In the second step, the predicted affinity was obtained for the sub-categories and the highest

sub-category was recommended. This study verified that a collaborative filtering method based on sub-categories leads to a high number of content selections by the target customers in comparison with collaborative filtering methods based only on the main category.



# I. 서론

최근 광고 시장이 급속도로 성장함에 따라, 고객이 요구하는 아이템을 빠르고 정확하게 제공해주는 것이 중요한 성공적 요인이 되고 있다. 이러한 개인의 요구에 맞는 콘텐츠를 제공해주는 것이 개인화 추천시스템이다. 광범위한 콘텐츠의 제공으로 콘텐츠 선택에 있어 개인화 추천에 많은 시간이 소요되고 있다. 이러한 문제점을 해결하기 위해 사용자가 원하는 콘텐츠를 예측하고, 빠른시간내에 콘텐츠를 선택할 수 있는 시스템이 중요시 될 것이다[1].

이러한 개인화의 흐름에 맞추어 개인화 추천 서비스가 제공되고 있는데, 개인화 추천 서비스란 구매 행동 분석을 토대로 사용자의 요구에 맞는 서비스 및 상품을 추천해주는 시스템이자[2], 사용자들이 선호하는 서비스

및 상품을 쉽게 찾을 수 있도록 도와주는 정보 필터링 기술이기도 하다.

모바일, 웹 콘텐츠 시장은 다양한 사용자 층의 확산으로 인해 미디어로서의 기본적인 가치 평가가 되고 있으며, 낮은 비용으로 광고주가 목표 고객에게 직접 접근할 수 있어 더욱 가치가 높게 평가되고 있다[7]. 광고 시장 규모가 커짐에 따라 경쟁도 치열해졌다. 경쟁이 치열해진 만큼 철저한 시장조사와 빠르고 정확한 정보의 습득을 바탕으로 사용자의 선호도에 따른 콘텐츠를 추천하는 시스템을 개발하는 일이 중요하다고 볼 수 있다.

본 연구에서는 콘텐츠를 추천하는데 있어 사용자가 원하는 콘텐츠를 추천하기 위한 방법으로 콘텐츠를 소분류로 세분화하여 콘텐츠를 추천하여 목표고객의 콘텐츠 선택의 횟수가 높은 시스템을 제시하고 실험을 통해 타당성을 검토한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구로 추천시스템과 적용기술을 소개하고 개인 추천 시스템에서 많이 사용되는 협업 필터링 기

법에 대하여 알아본다. 3단계에서는 협업 필터링의 핵심 기법에 대하여 서술한다. 4장에서는 제안한 시스템에 대한 실험 및 평가가 이루어지고, 5장에서는 결론 및 향후 연구 과제를 제시한다.



## II. 관련연구

본 장에서는 본 연구와 관련된 추천시스템과 협업 필터링의 개념 및 필요성에 대해 알아보고 협업 필터링 추천 방법의 문제점에 대해 살펴본다. 관련연구의 목적은 본 연구에서 제안한 카테고리 기반 협업 필터링 추천 시스템의 이론적 배경을 마련하고 이해하기 위함이다.

### 1. 추천 시스템(Recommendation System)

추천 시스템은 관심을 가지는 상품에 관한 정보나 인구통계학적 정보, 과거 구매행동을 토대로 고객의 요구에 맞는 상품을 추천해주는 시스템을 말한다[3]. 추천 시스템은 현재 Amazon.com, CDnow.com등의 전자

상거래 사이트에서 적용되고 있으며[7], 동영상 추천, 음악 추천, 도서 추천 등의 다양한 분야에서 이용이 되고 있는 추세이다. 콘텐츠의 증가로 인해 사용자가 원하는 콘텐츠를 선택하기 어려움짐에 따라 사용자가 요구하는 콘텐츠를 필터링하여 제공하는 시스템이 요구되고 있다. 개인화된 콘텐츠를 추천해주기 위해서는 사용자의 인구통계학적 속성 정보(Demographic features), 콘텐츠의 속성 정보(Contents Feature), 사용자의 과거 이용 정보(Usage History), 상품의 과거 구매 정보(Sales History)등이 고려 되어져야 한다.

그림 1은 추천 시스템에서 콘텐츠를 필터링하여 추천하는 내용을 나타낸다.

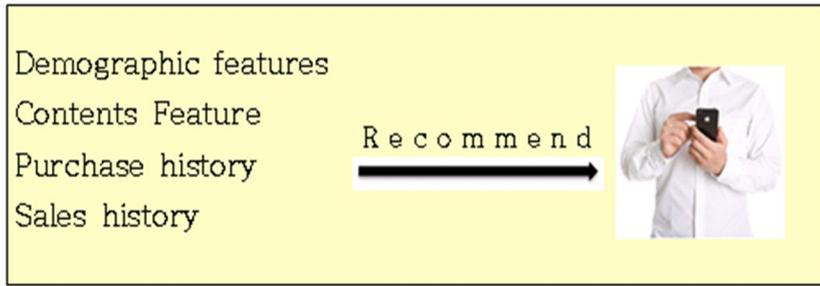


그림 1. 추천 시스템

개인화는 모바일 콘텐츠 시장에서 목표 고객이 요구하는 특정한 정보에 대해 빠르게 대응하여 정보를 제공할 수 있다는 것을 의미한다. 모바일 상에서의 개인화는 각 이용자의 성향과 요구를 만족시켜 이용자의 콘텐츠 사용에 신뢰도를 높여줄 뿐 아니라 일대일 마케팅을 가능하게 해준다는 점에서 큰 의의가 있다[9]. 또한 바쁘게 생활하는 현대인들이 제공받고자 하는 콘텐츠의 정보를 빠르게 검색할 수 있다는 점에서 개인화 추천 시스템의 중요성을 인식할 수 있다.

## 2. 추천 기술의 종류

구분	추천방식	개인화	추천효율
----	------	-----	------

개인화 추천기법으로 사용되는 추천 기술은 약간의 차이는 있으나 크게 3가지로 분류 할 수 있다. 첫 번째 규칙기반 필터링, 두 번째 협업 필터링, 세 번째 학습 에이전트 방법이다[4]. 세 가지 기술 외에도 내용기반 필터링 기술과 인구통계학적 필터링 기술이 있지만 대량의 데이터에서의 효율 저하로 개인화 추천 시스템에는 사용되기 어렵다고 알려져 있다.

위의 세가지 방법 중 개인화 추천시스템에 현재 가장 많이 사용 되어지고 있는 기법은 협업 필터링 기술이다.

		추천	
협업 필터링	다수의 사용자로부터 얻은 기호정보를 이용해 특정 고객의 관심사를 예측하여 상품 추천	가능	데이터 충분시 높은 효율
규칙기반 필터링	사용자 프로파일, 규칙을 이용해 개인화된 추천제공	가능	
학습 에이전트	웹사이트 방문기록, 횟수, 접속장소, 시간 등의 로그파일 분석이용	가능	
내용기반 필터링	시스템이 보유하고 있는 콘텐츠 속성 정보중, 단어를 입력했을 때 속성을 지닌 콘텐츠를 추천	어려움	대량의 데이터에서 효율 저하
인구통계학적 필터링	인구통계학적 정보 기반으로 유사한 이용자 참조해 항목 추천	불가능	

표 1은 추천 시스템에서 사용 되어지는 추천 기술의 종류를 나타낸 것

이다.

표 1. 추천 기술 종류

첫 번째, 규칙기반 필터링(rule-based recommendation)은 사용자에게 개인신상, 관심분야, 선호도 등을 질문 하여 사용자의 프로파일을 획득하게 되고 이러한 사용자의 정보를 기반으로 고객의 심리와 선호도 정보에 알맞은 정보 및 상품을 추천하는 기법이다. 규칙은 전문가에 의해 직접 작성되거나 시스템에 의해 생성될 수 있다. 두 번째, 학습 에이전트(Learning agent)는 웹 사이트 방문기록 및 횟수, 접속장소, 시간 등 일종의 로그 파일 분석을 통해 사용자의 속성, 습관, 개인의 선호를 추적하는 학습 에이전트를 이용하는 개인화 기법이다[9]. 세 번째, 협업 필터링(Collaborative Filtering)은 많은 사용자들로부터 얻은 기호정보(Taste Information)에 따라 고객과 유사한 성향을 갖는 이웃 고객들을 선별하고 이들이 공통적으로 선호하는 아이템을 추천하는 방법이다.

### 3. 협업 필터링(Collaborative Filtering) 기법

개인화 추천 시스템에 사용되는 여러 기법들 중 전자상거래 분야에서 성공적으로 적용된 기법은 협업 필터링(Collaborative Filtering)이다

[5].

본 논문에서 추천시스템을 설계 하는 데 사용된 협업 필터링 기법은 다음과 같다.

#### 가. 협업 필터링 추천 기법

유사도가 높은 사용자들간 협업 필터링 추천 기법은 그림2와 같이 나타낼 수 있다.

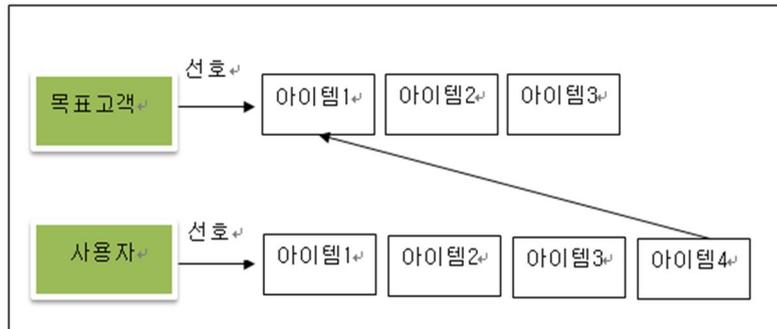


그림 2. 협업 필터링 추천 기법

개인화 추천 시스템은 구매 행동 분석을 토대로 사용자의 요구에 맞는 서비스 및 상품을 추천해주는 시스템으로 사용자들이 선호하는 서비스 및 상품을 쉽게 찾을 수 있도록 도와주는 정보 필터링 기술이다. 즉, 다수의 사용자로부터 얻은 기호 정보에 따라 목표고객의 관심사를 예측하여 상품을 추천하는 기법을 의미한다[9]. 과거에 자신이 좋아했던 상품과 비슷한 상품이면 좋아하는 경향이 있고 반대로 싫어했던 상품과 비슷한 상품이면 싫어하는 경향이 있다는 점을 기반으로 하고 있다. 예를 들어 1명의 목표고객과 4명의 사용자가 4개의 제품에 대해 1에서 5점 사이의

점수로 자신이 선호하는 제품을 평가(rating)했다고 가정하였을 경우 표2와 같이 C사용자가 목표고객과 더 가까운 취향을 가지고 있다는 것을 예측할 수 있다. 따라서 목표고객이 아직 평가하지 않은 제품을 A, B, D의 사용자보다 C사용자를 참조하는 것이 더 낫다고 판단할 수 있는데, 이와 같은 방법에 기초한 것이 협업 필터링 방법이다.

표 2. 협업 필터링 추천의 예

제품 평가자	도브	미장센	궁	엘라스틴
목표고객	5	4	3	평가정보없음
A	1	2	3	2
B	2	1	5	3
C	4	5	3	1
D	5	1	1	3

표 2는 목표고객에게 상품을 추천하기 위한 협업 필터링 추천의 예를

나타낸 표이다.

비슷한 취향을 가진 고객들에게 서로 아직 구매하지 않은 상품들은 교차 추천하거나 분류된 고객의 취향이나 생활 형태에 따라 관련 상품을 추천하는 형태의 서비스를 제공하기 위해 사용된다.

#### 나. 협업 필터링 종류

협업 필터링 기법은 사용자 기반 협업 필터링(user-based collaborative filtering)과 아이템 기반 협업 필터링(item-based collaborative filtering)으로 나눌 수 있다[6].

사용자 기반 협업 필터링(user-based collaborative filtering)은 사용자가 입력한 선호도 정보를 가지고, 고객과 유사한 성향을 갖는 이웃을 선별하여 공통적으로 선호하는 아이템을 추천하는 방법[7]으로 공통적으로 선호하는 아이템이 없으면 추천을 할 수 없는 단점이 있다.

아이템 기반 협업 필터링(item-based collaborative filtering)은 사용자들이 과거에 좋아했던 아이템들과 비슷한 아이템은 좋아하고 싫어했던 아이템과 비슷한 아이템은 싫어한다는 점을 기반으로 데이터를 분석하여 선호도를 예측[10]하는 기법이다.

유사도 높은 사용자를 선정하는 방법은 보편적으로 K-최근접 이웃(K Nearest Neighbor, KNN)이 사용되고 있다[10]. 고객과 이웃들간의 유사도를 구해 가장 높은 유사도(similarity)를 가지는 k개의 최근접 이웃을 구성하여, 가중치를 이용하여 선호도 예측 값을 계산한다.

본 논문에서는 사용자 기반 협업 필터링과 아이템 기반 협업 필터링의 단점을 보완하여 카테고리기반 협업 필터링 기법을 적용하여 보다 빠르게 목표고객이 원하는 정보를 제공하기 위한 방법으로 카테고리 세분화 협업 필터링 기법을 제안하였다.

## 다. 사용자 기반 협업 필터링 과정

그림 3은 협업 필터링 기법으로 콘텐츠를 추천하는 과정을 나타낸 것이며, 다음의 단계를 수행한다.



그림 3. 협업 필터링 추천 과정

1단계 : 아이템에 대한 평가치 매트릭스 준비

목표고객과 사용자간의 아이템 평가치 데이터를 준비한다.

## 2단계 : 최근접 이웃 구성

목표 고객과 다른 고객들간의 유사도를 계산한다. 유사도 계산식으로는 피어슨 상관계수 (Pearson correlation coefficient)와 코사인 유사도(cosine similarity)가 주로 사용되며, 수치(Rating)의 의미에 따라 피어슨 상관계수를 사용할지 코사인 방법을 사용할 것인지를 결정한다. 선호 또는 비선호의 의미를 포함할 경우 피어슨 상관계수 공식을 사용한다. 예를 들어 Rating이 1에서 5까지의 만족도 값을 의미할 경우 적절하다. 코사인 유사도 공식은 Rating값이 선호의 의미만을 포함할 경우 사용한다. 피어슨 상관계수는 두 개의 속성간에 상관성을 검사하여 1이면 같은성향, 0이면 상관없음, -1이면 반대성향을 나타낸다.

$$r_{xy} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

수식 1. 피어슨 상관계수 공식

여기서  $x$ 와  $y$ 는 고객  $x$ ,  $y$ 가 공통으로 평가한 아이템  $i$ 의 평가치를 의미한다. 그리고  $\bar{x}$ ,  $\bar{y}$ 는 고객  $x$ ,  $y$ 의 이용 가능한 평가치들의 평균값을 뜻한다.

코사인 유사도 공식은 두 벡터의 각도만을 고려하므로 벡터의 크기는 무시될 수 있기 때문에 선호도 정보가 없고 콘텐츠 이용 횟수나 구매 횟수 정보만 있을 경우 사용된다. 1이면 같은성향, 0이면 다른 성향을 나타낸다.

$$\cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{|\vec{i}| * |\vec{j}|}$$

### 수식 2. 코사인 유사도 공식

본 논문에서는 피어슨 상관계수(Pearson Correlation Coefficient)와 코사인(Cosine) 방법 중 여러 연구에서 성과가 높다고 알려진 피어슨 상관계수 방법을 이용하여 사용자와 목표고객간의 유사도 계산에 사용하였다.

### 3단계 : 선호도 예측 값 계산

앞에서 계산된 목표 고객과 다른 사용자들간의 유사도를 가중치로 하여 각 상품에 대한 선호도 예측 값을 계산한다. 선호도 예측 식은 다음과 같다.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) \times w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

수식 3. 선호도 예측 공식

## 라. 아이템 기반 협업 필터링 기법

아이템기반 협업 필터링(item-based Collaborative Filtering)은 공통으로 평가한 아이템의 평가 정보를 활용하여 유사 선호를 가진 사용자를 찾아 추천해 주는 기법이다[6].

구현방법은 목표고객과 같은 아이템을 평가한 사용자들을 선별하고, 선별한 사용자들이 선호도를 입력한 아이템간 유사도를 계산한다. 유사도 계산방법은 사용자 기반 협업 필터링과 동일하다. 계산된 유사도를 가중치로 사용하여 각 상품에 대한 선호도 예측 값을 계산한다. 선호도 예측 식은 수식5와 같다.

$$P_{u,i} = \frac{\sum_{allSimilarItems,N} (S_{i,N} * R_{u,N})}{\sum_{allSimilarItems,N} (|S_{i,N}|)}$$

수식 4. 아이템 기반 선호도 예측 공식

이 방법은 고객 간의 유사도가 고려되지 않기 때문에 특정 고객과 선호도가 다른 사용자들의 평가가 기반이 되면 아이템간의 상관관계 정확도가 떨어지는 한계점을 가지고 있다[9].



#### 4. 협업 필터링의 문제점

대표적인 협업 필터링의 한계점은 희박성(Sparsity), 확장성(Scalability), 투명성(Transparency)이다[11].

사용자 기반 협업 필터링의 경우 사용자와 아이템의 수가 많아지면 목표고객과 사용자간의 최근접 이웃을 구성하기 위한 연산이 증대되어 확장성(Scalability), 선호도가 입력되지 않은 상품의 수가 많아짐에 따라 신뢰성이 다소 낮아지는 희박성의 문제점이 있다[12]. 아이템 기반 협업 필터링에서는 데이터 희박성(Sparsity) 문제는 없지만, 사용자간의 개인 취향을 반영 하지 못하는 투명성(Transparency)의 문제점이 있다.

### III. 설계 및 구현

본 연구에서는 목표고객에게 원하는 콘텐츠를 추천하기 위한 방법으로 카테고리 대분류로 추천하는 방법보다, 카테고리 소분류로 추천하였을 경우 목표고객이 원하는 콘텐츠를 선택하는 횟수가 높은 시스템을 제안한다.

#### 1. 카테고리 세분화 협업 필터링 추천시스템 구성

본 논문에서는 광범위하고 다양한 콘텐츠 중에서 사용자가 원하는 콘텐츠 선택에 도움을 주기 위하여 카테고리 세분화 협업 필터링 추천 기법을 이용하여 목표고객이 원하는 콘텐츠를 좀 더 빠른 시간 내에 찾을 수 있도록 상위카테고리 아래에 소분류 3개의 카테고리로 세분화 하여 목표

고객에게 콘텐츠를 추천하여 대분류 카테고리 추천방법보다 소분류 카테고리 추천방법으로 목표고객의 선택 횟수를 높일 수 있는 시스템을 제안한다. 그림 4는 카테고리 세분화 협업 필터링의 과정을 나타낸 그림이다.



그림 4. 추천시스템 구조

사용자와 목표고객이 광고를 보면 클릭한 횟수가 저장된다. 각 광고는 대분류/소분류로 분류된 카테고리에 저장되어 있기 때문에 카테고리의 사용자 선호도 정보를 업데이트 한다.(선호도 정보는 광고를 클릭한 횟수를

의미함)

통합된 데이터베이스에서 카테고리 세분화 협업 필터링 기법을 사용하여 가장 선호도 높은 광고를 추천한다.

본 장에서는 카테고리 세분화 협업 필터링의 단계별 과정을 설명하고, 설계된 추천 시스템을 적용하여 4장에서 실험을 바탕으로 검증을 하는 내용을 다룬다. 피부과 진료항목의 콘텐츠를 선택할 수 있는 시스템으로 목표고객과 사용자들의 진료항목 콘텐츠의 선택 횟수를 이용한다.

카테고리를 세분화하여 소분류간 유사도를 계산하여 소분류 선호도 예측 값을 계산하여 선호도 예측 값이 가장 높은 소분류의 콘텐츠를 추천하게 된다.

## 2. 카테고리 세분화 협업 필터링 단계

카테고리 세분화 협업 필터링 추천은 다음의 수행 단계로 진행된다.

1 단계 : 목표고객과 사용자 평가치 매트릭스 준비

목표고객과 사용자 20 명이 선택한 컨텐츠 선택 정보 매트릭스를

수집한다.



2 단계 : 유사도 계산하여 최근접 이웃 구성

**목표고객과 사용자 20명간의 유사도를 계산하여 유사도가 높은  
    최상위 5명의 최근접 이웃을 구성한다. 유사도 계산은 피어슨 상관계수  
    공식을 사용한다.**

3 단계 - 대분류 카테고리간 유사도 계산

**최근접 이웃 5명의 이웃들이 본 대분류 카테고리 데이터를 이용하여  
    목표고객과의 유사도를 계산하여 테이블화 한다.**

4 단계 - 대분류 선호도 예측값 계산

**계산된 유사도 값을 가중치로 사용하여 목표 고객과 각 카테고리간  
    선호도 예측 값을 계산한다.**

**5 단계 - 소분류 카테고리 평가치 매트릭스 준비**

선호도 예측값이 가장 높은 대분류 카테고리를 선정하여 예측값이 가장 높은 대분류에 속하는 목표고객, 사용자의 소분류 카테고리의 선택정보를 준비한다.

**6 단계 - 소분류 카테고리간 유사도 계산**

목표고객과 이웃고객들의 소분류 카테고리간 유사도를 계산한다.

**7 단계 - 소분류 선호도 예측 값 계산**

계산된 유사도 테이블 각각의 유사도 값을 가중치로 사용하여 목표 고객에게 추천하고자 하는 콘텐츠 카테고리에 대한 선호도 예측 값을 계산하여 가장 높은 소분류 카테고리를 추천한다.

## IV. 실험 및 성능 평가

사용자와 목표고객 콘텐츠 선택 Data를 이용하여 대분류 카테고리 기반 협업 필터링의 콘텐츠 추천 방법과 소분류 카테고리 기반 협업 필터링의 콘텐츠 추천 방법을 구현해보고, 목표 고객이 콘텐츠를 선택하는 횟수를 비교하여 평가한다.

먼저, 목표고객과 20명의 사용자들에게 대분류, 소분류로 분류된 카테고리 정보를 표 3과 같이 보여준다.

**표 3. 콘텐츠 대분류/ 소분류 구분 정보**

대분류 카테고리		소분류 카테고리	
대분류1	피부	소분류1 소분류2 소분류3	•레이저토닝 •여드름치료 •보톡스
대분류2	비만관리	소분류1 소분류2 소분류3	•카복시테라피 •메조테라피 •지방분해주사
대분류3	흉터치료	소분류1 소분류2 소분류3	•튼살치료 •재생관리 •흉터치료
대분류4	여드름	소분류1 소분류2 소분류3	•블랙필 •여드름 짜기 •모공관리
대분류5	제모	소분류1 소분류2 소분류3	•팔 •다리 •겨드랑이

## 1. 카테고리 세분화 협업 필터링 추천 실험

카테고리 세분화 협업 필터링 추천 실험은 다음의 단계로 진행된다.

1단계 : 목표고객과 사용자 20명의 콘텐츠 선택 정보 준비

대분류 카테고리 데이터를 목표고객과 사용자들에게 제공하고 목표고

객에게 모바일 콘텐츠를 추천하기 위하여 5개의 카테고리에 대한 목표고

객과 20명의 고객들의 평가치를 가져온다.

표 4. 목표고객 평가치

대분류 사용자	대분류1	대분류2	대분류3	대분류4	대분류5
Target	7	0	5	4	0

표 5. 사용자 20명 평가치

대분류 사용자	대분류1	대분류2	대분류3	대분류4	대분류5
A	7	1	3	0	5
B	3	0	1	0	1
C	5	0	2	1	3
D	4	1	2	0	6
E	2	0	5	2	7
:					
R	6	0	0	1	4
S	0	4	6	0	1
T	3	1	4	1	3

표 4, 표 5는 목표고객과, 사용자들이 콘텐츠를 선택 했을 때 각 카테고리 별로 누적된 데이터 횟수를 나타낸다.

2 단계 : 최근접 이웃 구성

목표고객과 다른 사용자들의 유사도를 피어슨 상관계수 방법을 사용하여 구하고, K-NN기법으로 표 6과 같이 최상위 5명으로 최근접 이웃을 구성한다.

표 6. 목표고객과 사용자간 유사도

사용자	유사도
A	0.78
B	0.66
I	0.82
O	0.79
R	0.73

3단계 : 대분류 카테고리 유사도 계산

K-NN기법으로 선정된 5명의 이웃들이 본 대분류 카테고리를 대상으로 대분류 카테고리간의 유사도를 계산한다. 표7은 목표고객과 최근접 이웃으로 선정된 5명과의 대분류 카테고리간 유사도를 나타낸다.

표 7. 대분류 카테고리간 유사도

목표고객 사용자	대분류1	대분류2	대분류3	대분류4	대분류5
대분류1	1	0	0.63	0.33	0.05
대분류2	0	1	0.07	-0.08	0.67
대분류3	0.63	0.07	1	0.62	-0.46
대분류4	0.33	-0.08	0.62	1	-0.5
대분류5	0.05	0.67	-0.46	-0.5	1

4단계 : 대분류 선호도 예측 값 계산

대분류 카테고리에 대한 선호도 예측 값을 계산하여 값이 가장 높은

대분류 카테고리를 추천한다.

표 8. 대분류 선호도 예측 값

대분류 사용자	대분류1	대분류2	대분류3	대분류4	대분류5
Target	7	0.5	5	4	3.21

표 8과 같이 목표고객이 선택하지 않은 대분류 중 선호도 예측 값이 높은 대분류5에 해당하는 콘텐츠를 추천한다.

5단계 : 소분류 카테고리 정보 준비

선호도 예측값이 가장 높은 대분류 카테고리5에 해당하는 목표고객, 사용자의 소분류 카테고리의 선택 정보를 그림5, 그림6과 같이 준비한다.

	대분류1			대분류2			대분류3			대분류4			대분류5		
	소분류 1	소분류 2	소분류 3	소분류 1	소분류 2	소분류 3	소분류 1	소분류 2	소분류 3	소분류 1	소분류 2	소분류 3	소분류 1	소분류 2	소분류 3
Target	6	0	3	0	0	0	2	4	5	4	5	5	8	0	0

그림 5. 목표고객 소분류 카테고리 선택 정보

	대분류1			대분류2			대분류3			대분류4			대분류5		
	소분류 1	소분류 2	소분류 3	소분류 1	소분류 2	소분류 3	소분류 1	소분류 2	소분류 3	소분류 1	소분류 2	소분류 3	소분류 1	소분류 2	소분류 3
A	10	4	1	3	0	0	0	5	2	0	0	0	5	0	3
B	1	2	6	0	0	0	4	0	1	0	0	0	6	0	0
C	4	0	1	0	0	0	2	1	6	1	2	5	1	6	0
D	1	0	4	1	5	2	0	7	0	0	0	0	4	6	3
E	5	0	1	0	0	0	0	5	1	3	1	4	3	1	2
F	0	0	0	0	0	0	2	4	8	2	1	0	2	5	0
G	0	0	0	2	7	5	0	2	0	1	1	4	1	4	1
Q	9	0	4	5	3	0	0	0	0	3	2	4	6	5	3
R	8	3	5	0	0	0	0	0	0	0	0	0	4	2	5
S	0	0	0	7	3	3	3	5	4	0	0	0	0	8	3
T	8	5	5	4	2	5	3	8	1	1	5	2	5	8	3

그림 6. 사용자 소분류 카테고리 선택 정보

6 단계 : 소분류 카테고리간 유사도

이웃 고객들이 본 대분류 5에 해당하는 소분류 카테고리와의 목표고객이 본 소분류 카테고리를 대상으로 소분류 카테고리간 유사도를 표 9와 같이 계산하여 나타낸다.

표 9. 소분류 카테고리간 유사도

목표고객 사용자	소분류1	소분류2	소분류3
소분류1	1	-0.36	-0.78
소분류2	-0.36	1	0.83
소분류3	-0.78	0.83	1

7 단계 : 소분류 선호도 예측 값 계산

계산된 유사도 테이블 각각의 유사도 값을 가중치로 사용하여 목표 고객에게 추천하고자 하는 콘텐츠 카테고리에 대한 선호도 예측 값을

계산하여 가장 높은 소분류 카테고리를 추천한다.



표 10. 소분류 선호도 예측 값

	소분류1	소분류2	소분류3
Target	1	6.3	0.34

표 10과 같이 목표 고객에게 소분류 2에 해당하는 콘텐츠를 추천함을 알 수 있다.

## 2. 실험 평가

목표고객과 사용자 20명의 콘텐츠 선호도를 이용하여 목표고객이 원하는 콘텐츠를 빠르게 검색할 수 있는 시스템으로 카테고리 세분화 협업 필터링 기법을 적용하여 추천시스템을 제안하였다. 피부과 진료 항목 광고 150개를 8일간 대분류, 소분류로 나누어 실행하였다.

목표고객이 콘텐츠를 선택하는 횟수가 일자가 경과함에 따라 대분류

카테고리로 추천한 경우와, 소분류로 세분화하여 추천 하였을 경우 어떻게  
변화 하는지 관찰 하였다.

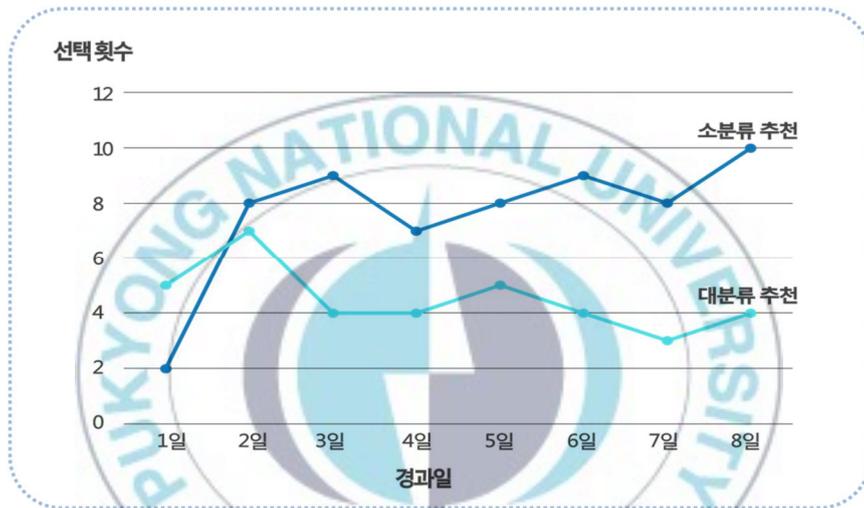
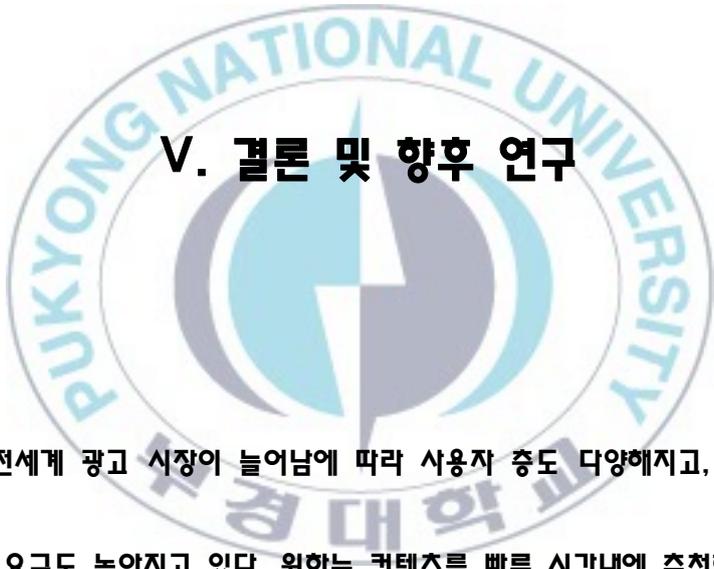


그림 7. 날짜경과에 따른 목표고객 선택 횟수

그림 7은 협업 필터링 기법을 이용한 대분류 콘텐츠 추천 방법과,  
소분류 콘텐츠 추천 방법을 적용했을 때 목표고객의 선택 값의 변화를  
그래프로 표현한 것이다. 소분류 카테고리 기반 협업 필터링 기법의 추천  
방법이 대분류 카테고리 기반 협업 필터링 기법으로 콘텐츠 추천을 진행한

경우보다, 목표고객에게 추천 되는 콘텐츠의 추천 범위가 좁아지므로  
일자가 지남에 따라 목표고객이 콘텐츠를 선택하는 횟수가 높게  
나타난다는 것을 알 수 있다.



## V. 결론 및 향후 연구

전세계 광고 시장이 늘어남에 따라 사용자 층도 다양해지고, 사용자  
의 요구도 높아지고 있다. 원하는 콘텐츠를 빠른 시간내에 추천하여 선  
택하기 위해서는 사용자에게 다양하고 광범위한 콘텐츠를 추천하는 것  
보다 목표고객이 요구하는 콘텐츠를 파악하여 추천해주는 시스템이 중  
요한 역할을 할 것이다.

본 논문에서는 카테고리 협업 필터링 기반의 피부과 진료과목 추천과정에서 기존의 대분류 카테고리 기반 협업 필터링의 추천방법의 문제점 중 추천 콘텐츠 범위가 넓어 원하는 콘텐츠를 선택하는데 있어 시간이 많이 걸리는 문제점을 해결하고자, 소분류 카테고리 기반의 협업 필터링 방법을 제안하고 설계하였다. 실험에 의하면 목표고객과 사용시간 유사도가 높은 최근접 이웃을 구성하여 선호도 예측 값이 가장 높은 대분류 카테고리 콘텐츠 추천하였을 경우보다, 소분류로 카테고리를 세분화하여 콘텐츠를 추천하였을 때 목표고객이 원하는 콘텐츠를 선택하는 횟수가 높았다는 것을 확인하였다. 대분류로 콘텐츠를 추천하는 경우 많은 양의 콘텐츠에서 선택하는데 오랜시간이 걸리는데 비해, 소분류로 카테고리를 세분화하여 콘텐츠를 추천하는 경우 목표고객이 원하는 콘텐츠의 범위가 좁아져 선택하는데 있어 편리하고 선택 횟수도 높았다.

향후 연구 과제로는 제시한 대분류, 소분류의 분류 방법을 보다 다양

한 분야에 적용하고 정확성을 높이기 위한 방법에 대한 연구가 필요할 것이다.



## 참고 문헌

- [1] Pottoniee, O., "Ubiquitous Personalization: a Smart Card Based Approach, Gemplus Research Labs, BP 100, 138811 Gemenos Cedex, 2000.
- [2] Mulvenna, M., Sarabjot, S. A. and Alex, G. B., " Personalization on the Net Using Web Mining; Interoduction", Communications of the ACM, Vol. 43, No. 8, 2000.
- [3] Sarwar, B. M., B. Karypis, J.A. Konstan, and J. Riedl. "Item-Based Collaborative Filtering Recommen-dation Algorithms,"Proceedings of the 10<sup>th</sup> Inter-national World Wide Web Conference. 2001.
- [4] Kuo, Y.-F., and L.-S. Chen,. "Personalization Tech-nology

Application to Internet Content Provider,” Expert Systems with Applications, Vol. 21, No. 4, 2001.

[5] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.,

“Grouplens: an open architecture for collaborative filtering of netnews,” Proceedings of the ACM Conference on Computer Supported Cooperative Work, 1994.

[6] M. Deshpande & G. Karypis, Item-Based Top-N Recommendation Algorithms, ACM Trans. Information Systems, vol.22, no.1, 2004.

[7] 김은규, 조인휘, “카테고리 협업 필터링 기반의 모바일 콘텐츠 추천 시스템 설계”, 한양대학교, 2012.

[8] 최현석, 배효철, 서정진, 윤경로, ‘감성 및 상황 정보 융합 기반의 확장

된 협업 필터링 기법을 이용한 음악추천시스템”, 한국방송공학회 학술 발표대회, 2011.

[9] 강민정, “스마트폰 이용자의 어플리케이션 구매를 위한 추천 시스템”, 홍익대학교 2010.

[10] 김청송, 강재우, “추천 시스템에서의 협동적 필터링에 관한 연구 및 분석”, 고려대학교, 2010.

[11] 박석두, “추천시스템의 성능 향상을 위한 단계적 협업 필터링의 설계 및 구현”, 아주대학교, 2007.

[12] 김재경, 안도현, 조운호, “개인별 상품추천 시스템, Web CF-PT 웹마 이닝과 상품계층도를 이용한 협업 필터링”, 경영정보학연구, 제 15 권 제1호, 2005.

## 감사의 글

남들보다 한 학기 더 길었던 대학원 생활의 마감을 앞둔 지금, 많은 추억들이 스쳐 지나갑니다. 대학원 생활은 저에게 좋은 인연과 다양한 인생을 경험하게 해 줄 수 있는 기회가 되었습니다.

육아와 회사생활과 병행하며 논문을 작성하느라 늘 정신없는 저에게 인자하게 하나하나 지적해주시고 가르쳐 주시느라 정말 고생 많이하셨던 윤성대 교수님 진심으로 감사드립니다. 교수님 덕분에 무사히 졸업을 할 수 있게 되었습니다. 가르침 늘 가슴속 깊이 간직하겠습니다. 논문준비를 위해 세심하게 신경 많이 써주시고, 조언 및 좋은 말씀을 많이 해주셔서 큰 위로와 힘이 되어주신 김명희 교수님 감사합니다. 두 이쁜 공주님들을 두시고도 참 열심히 생활하고 노력하시는 교수님 모습 보면서 때로는 위로를 받고, 때로는 큰 힘이 생기기도 하였습니다. 따뜻한 말씀들로 인하여 힘들었던 대학원 생활에 큰 힘이 되었습니다. 미국에서 이쁜 두 공주님과 좋은 추억과 경험 많이 쌓고 오시길 바라겠습니다.

바쁘신 와중에도 논문 심사를 위해 세심한 관심을 보여주신 조유현 교수님, 서경룡 교수님 감사드립니다.

많은 수업이 있음에도 불구하고 학술대회 발표에서부터 논문을 꼼꼼히 봐주시며 검토해 주셨던 윤소영 선생님 감사합니다. 너무나 부족함이 많은 논문인데도 불구하고 세심하게 검토해주셔서 큰 도움이 되었습니다. 지금은 연구실에 안계시지만 많은 도움과, 좋은 말씀 주셨던 황순환 선생님 감사드립니다. 늘 건강하시고 하시는 일마다 행운이 깃들길 바랍니다.

같이 논문 준비를 하면서 도움 주시고, 좋은 말씀 해주신 임선자 선

생님 감사합니다. 늘 노력하시고 항상 배우는 자세를 본받아야겠다라는 생각을 많이 가지게 되었습니다. 박사논문 마무리 잘 하시길 바라겠습니다. 먼저 졸업을 하였던 대학원 동기 김주영 선생님, 이영미 선생님, 김동영 선생님, 허준호 선생님 다들 고생 많이 하셨고 감사합니다.

저의 대학원 생활로 인해 너무나 고생많이 하셨던 우리 부모님 감사합니다. 직장 다니느라 새벽에 나가서 밤늦게 들어오는 딸 덕분에 두 아이들과 사위 저녁까지 손수 차리느라 너무 고생많으셨던 우리 엄마 정말 감사합니다. 시간 나실 때 마다 집에 오셔서 아이들을 함께 봐주시고 묵묵히 뒤에서 지원 해주셨던 우리 아빠. 꼭 제가 하고 싶었던 꿈을 이루어 보답하겠습니다.

대학원 입학하였을 때 등록금까지 내어주시며 함께 기뻐해주시신 아버님, 어머님 감사합니다. 멀리 중국에 계신 아버님 항상 존경합니다. 항상 제가 하고자 하는 일을 든든하게 지원해주는 든든한 나의 신랑 너무 감사합니다. 내가 기뻐하면 나보다 더 기뻐하는 신랑 덕분에 대학원생활 문제없이 잘 할 수 있었고, 내가 미처 하지 못하는 일처리를 꼼꼼하게 뒤에서 다 처리해주는 우리 신랑. 항상 사랑하고 앞으로는 우리 가족을 위해 많이 노력하는 엄마, 아내가 되겠습니다.

제 논문은 혼자 힘으로 쓰여진 논문이 아님을 압니다. 많은 분들의 도움 덕분에 무사히 마무리를 할 수 있었습니다. 앞으로 이러한 감사함을 가슴깊이 새겨 더 노력하며 살아가도록 노력하겠습니다. 많은 도움을 주신 모든 분들께 다시 한 번 감사하다는 말씀 전합니다.

2013년 1월  
김 은 숙 드림