

협업 필터링을 이용한 개인화 재능 추천 시스템

캡스톤 디자인2 결과보고서

순천향대학교
SOON CHUN HYANG
UNIVERSITY



학	번	20114016		
이	름	이태수		
학	과	컴퓨터 소프트웨어공학과		
지도	교수님	이화민 교수님		
작	품	이	름	협업 필터링을 이용한 개인화 재능 추천 시스템
강	의	명	캡스톤 디자인2	
담당	교수님	이임영 교수님		

차 례

1. 서론-----	1
2. 관련연구-----	2
2.1. 개인화 추천 시스템-----	2
2.2. 데이터 마이닝-----	4
3. 협업적 필터링 기법-----	7
3.1. 사용자기반 협업적 필터링-----	7
3.2. 아이템기반 협업적 필터링-----	9
3.3. 협업적 필터링의 문제점-----	9
3.4. 협업적 필터링의 문제점 보완-----	11
4. 개인의 특성을 이용한 협업적 필터링 기반의 개인화 재능 추천 시스템의 설계-----	12
4.1. Model2 기반의 JSP환경의 서버-클라이언트 구조-----	12
4.2. 개인화 재능 추천 시스템의 구성도-----	14
4.3. 인접 이웃 그룹 형성 및 추천 목록 생성 과정-----	15
4.4. 희박성 문제의 해결-----	17
5. 개인의 특성을 이용한 협업적 필터링 기반의 개인화 재능 추천 시스템의 구현-----	18
5.1. 구현 환경-----	18
5.2. 개인화 재능 추천 시스템 구현-----	18

1. 서론

전자상거래 시장은 스마트 디바이스의 등장 이후 모바일상의 거래를 중심으로 새로운 시장으로 재편되고 있다. 유저가 데이터에 접할 수 있는 기회가 점점 늘어나고 이용 가능한 수집된 정보의 양 또한 기하급수적으로 늘어나 있는 상태이다. 이로 인해 기하급수적으로 수집된 정보로 인한 정보의 범람은 고객이 선택을 하기 어렵게 만들어 전자상거래 시장의 발전에 방해요인이 되고 있다. 최근 소셜 커머스와 같이 고객의 참여에 따라 가격을 할인 받고 고객들 간에 적극적으로 정보를 공유하는 전자상거래 사이트가 늘어나면서 전자상거래 시장은 새로운 마케팅 전략이 요구된다.

최근 관련 분야에서 고객의 행동과 패턴을 이해하고 이를 구매요인으로 활용하기 위해 자동화된 개인화 서비스를 적극 도입하는 것을 추진하고 있다. 전자상거래 사이트는 사용자들의 구매내역과 장바구니이력, 장바구니에 담은 상품정보를 수집하고 이를 기반으로 고객에게 새로운 상품을 추천하는데 활용하고 있다. 협업적 필터링은 이와 같은 추천 시스템에 가장 많이 사용되는 알고리즘으로 학문적 관심영역을 넘어 아마존, 넷 플릭스 등 다양한 기업의 온라인 사이트에 적용되고 있다.

협업적 필터링은 유사도를 기반으로 아이템을 추천하는 방식으로 적용 대상에 따라 사용자 기반의 협업적 필터링과 아이템 기반의 협업적 필터링으로 나뉜다.

사용자 기반의 협업적 필터링 기법은 성향이 비슷한 사용자는 같은 아이템에 대해 비슷한 선호도를 가진다는 개념을 기반으로 한 추천 알고리즘이다. 비슷한 선호도를 가지는 사용자 그룹을 추출하고 추출된 그룹의 선호도를 바탕으로 해당 사용자의 특정 아이템에 대한 선호도를 예측한다. 아이템기반 협업적 필터링 기법은 사용자가 이전에 구매했던 과거의 아이템들과 비슷한 아이템을 구매할 가능성이 높다는 개념을 기반으로 한 추천알고리즘이다. 즉, 사용자는 유사한 아이템에 대해 유사한 선호도를 가질 가능성이 높다는 것을 기반으로 한다. 아이템기반 협업적 필터링에서는 사용자가 기존에 리뷰 등으로 평가한 각각의 상품들과 선호도를 예측하고자 하는 상품의 상관관계를 이용해서 선호도를 예측한다.

협업적 필터링은 다양한 비즈니스 분야에서 많은 유용성을 인정받고 있다. 하지만 데이터의 희소성, 확장성 처리 이슈는 중요한 문제점으로 제기되고 있으며 특히 희소성의 문제가 대두되고 있다. 따라서 본 연구에선 협업적 필터링 기반의 추천시스템에 존재하는 희소성 문제를 해결하기 위해 개인의 특성을 이용한 인구통계학적 기반 필터링을 함께 적용하여 성능을 향상시키며 향상된 알고리즘이 적용된 개인화 재능추천시스템을 설계하고 구현한다.

2. 관련 연구

본 장에서는 개인의 성향과 협업적 필터링을 이용한 개인화 재능추천 시스템을 설계하고 구현하는데 필요한 개인화 추천시스템, 데이터마이닝 등에 대해서 설명한다.

2.1. 개인화 추천 시스템

추천시스템은 고객이 관심을 가지는 아이템에 관련된 정보나 인구 통계학적 정보, 과거의 구매이력 분석을 기반으로 사용자의 요구에 맞는 아이템을 추천해주는 시스템을 말한다. 전자상거래 환경에서 점점 개인화 서비스의 중요성이 대두되고 있으며 개인화 서비스가 필요한 이유는 사용자의 입장에서 아이템을 찾는 데 들이는 시간적 노력을 감소시킬 수 있고 기업의 입장에서 적절한 아이템을 추천함으로써 사용자의 구매활동에 대한 참여를 높여 기업의 이익과 고객과의 유대감을 높일 수 있기 때문이다.

개인화는 주로 웹 환경에서 이뤄지며 이것은 웹이 사용자가 서비스 및 아이템을 이용하면서 발생하는 사용자의 데이터를 수집하고 분석하는데 최적의 환경이기 때문이다. 또한 기업의 입장에서 비용이 적게 소비되는 것과 현실세계에서 개개의 사용자에게 맞춤형 개인화를 제공하는 것에 비해 효율적인 제공이 가능하기 때문이다.

예를 들어 현실세계에서 개개의 사용자에게 맞춤형 개인화를 진행하면 사용자는 서비스를 받기 위해 서비스를 이용하는 지점이 위치한 곳에 방문하거나 서비스를 제공하는 곳에 전화나 메일을 통해 요청해야 하는 불편함이 생긴다. 반면 웹상에선 사용자가 원하는 시간에 웹에 접근만 가능하다면 아무런 제약 없이 개인화 서비스를 이용할 수 있다. 이러한 이유로 대부분의 개인화 서비스가 웹 사이트상에서 제공된다.

추천시스템을 위한 개인화 방법은 다음과 같다.

(1) 내용기반 필터링(content-based filtering)

내용기반 필터링 기법은 정보 검색이나 정보 필터링 연구에서 자연적으로 발전한 것으로 항목의 정보를 분석하여 사용자에게 항목을 추천하는 기법이다. 항목이 가지는 정보가 풍부하고 분석하기 쉬운 뉴스, 웹 페이지, 텍스트 등의 항목을 추천하는 시스템에 적합한 기법이다. 사용자 간의 선호도 정보를 이용하는 것이 아니라 각 사용자 별 특정 아이템 또는 아이템의 속성을 기반으로 필터링한다.

내용기반 필터링 기법의 동작은 상품의 추천을 위해 상품 또는 상품의 내용과 사용자가 요구하는 정보 간의 유사도를 계산하고 그 계산 결과를 목차형식으로 나타내는 것이다. 이것의 구현을 위해선 가중치 기법, 적합성, 피드백, 확률검색 모형 등의 방법을 사용한다.

내용기반 필터링 기법의 단점으로는 영화나 음악 같은 분야에서는 내용 분석이 정확하게 이뤄지기 어렵다는 것이다. 이러한 분야에서는 특징을 추출하기 어렵기 때문이다. 이러한 분야에 내용기반 필터링 기법을 적용하기 위해서는 각 아이템들의 내용을 일일이 정의하고 각 내용을 속성별로 정리하는 쉽지 않은 작업이 필요하다. 또한 내용기반 필터링 기법은 사용자의 사용자 프로파일과 비교하여 높은 점수를 갖는 아이템을 추천하기 때문에 유사한 아이템이 계속해서 추천되는 단점이 존재한다.

(2) 규칙기반 필터링(rule-based-filtering)

사용자에게 관심 분야나 다양한 형태의 선호도 등을 질의하여 그 사용자의 정보 프로파일을 획득하는 기법이다. 규칙기반 필터링의 동작은 취미와 특정 사항에 대한 선호도 등을 질문하고 답을 얻는 것을 통해서 그 사람의 프로파일을 획득한다. 이러한 데이터를 기반으로 사용자의 심리 정보와 사용자가 갖는 선호도 정보에 알맞은 정보 및 상품을 추천 또는 제공하는 방법이다.

사용자들의 선택에 의한 내용 전달이라는 측면에서 보면 웹 고객화(Customization)와 차이가 없는 것처럼 보인다. 웹 고객화란 적응적 웹 기술(Adaptive Web Technology)의 한 영역으로 고객화는 고객이 웹의 내용 구성이나 형태를 선택할 수 있는 것을 의미한다. 예를 들어 정보의 항목이나 기능 색 배치 등을 원하는 스타일로 할 수 있도록 제공되는 것이다. 그러나 규칙 기반 필터링에서 제공하는 질의는, 사용자들이 자신의 선택을 통해 내용을 스스로 구성하게 하기 위한 것이 그 목적이 아니라, 질의를 통해 사용자들을 구분하기 위한 목적으로 사용된다는 점에서 웹 개인화와는 큰 차이점이 있다.

(3) 인구통계학적기반 필터링(demographic-based filtering)

나이, 혈액형, 전공, 성별, 생활 수준, 거주지, 직업 등의 다양한 사용자의 개인 특성을 활용하여 추천을 수행하는 기법이다. 인구 통계학적 속성은 다양한 종류의 항목이나 항목 범주에 대한 사용자의 성향을 분석하기 쉬우며 사용자의 선호도, 장바구니 이력, 구매 이력과 같은 데이터가 없는 초기 단계의 추천시스템에서 사용자의 개인 성향을 분석하여 개인화된 추천기법을 적용하는 것이 가능하다.

(4) 협업적 필터링(collaborative filtering)

협업 필터링(collaborative filtering)은 많은 사용자들로부터 얻은 기호정보(taste information)에 따라 사용자들의 관심사들을 자동적으로 예측하게 해주는 방법이다. 협력 필터링 접근법의 근본적인 가정은 사용자들의 과거의 경향이 미래에서도 그대로 유지 될 것이라는 전제에 있다. 예를 들어, 음악에 관한 협력 필터링 혹은 추천시스템(recommendation system)은 사용자들의 기호(좋음, 싫음)에 대한 부분적인 목록(partial list)을 이용하여 그 사용자의 음악에 대한 기호를 예측하게 된다. 이 시스템은 특정 사용자의 정보에만 국한 된 것이 아니라 많은 사용자들로부터 수집한 정보를 사용한다는 것이 특징이다. 이것이 단순히 투표를 한 수를 기반으로 각 아이템의 관심사에 대한 평균적인 평가로 처리하는 방법과 차별화 된 것이다. 즉 고객들의 선호도와 관심 표현을 바탕으로 선호도, 관심에서 비슷한 패턴을 가진 고객들을 식별해 내는 기법이다. 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.

협업적 필터링 기법은 유사도를 계산하는 대상이 누구냐에 따라 사용자 기반 협업적 필터링(User-based Collaborative Filtering)과 아이템 기반 협업적 필터링(Item-based Collaborative Filtering)으로 나뉜다.

2.2. 데이터마이닝

데이터 마이닝(data mining)은 대규모로 저장된 데이터 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 찾아 내는 것이다. 다른 말로는 KDD(데이터베이스 속의 지식 발견, knowledge-discovery in databases)라고도 일컫는다.

데이터 마이닝은 통계학에서 패턴 인식에 이르는 다양한 계량 기법을 사용한다. 데이터 마이닝 기법은 통계학 쪽에서 발전한 탐색적 자료 분석, 가설 검정, 다변량 분석, 시계열 분석, 일반선형모형 등의 방법론과 데이터베이스 쪽에서 발전한 OLAP (온라인 분석 처리:On-Line Analytic Processing), 인공지능 진영에서 발전한 SOM, 신경망, 전문가 시스템 등의 기술적인 방법론이 쓰인다.

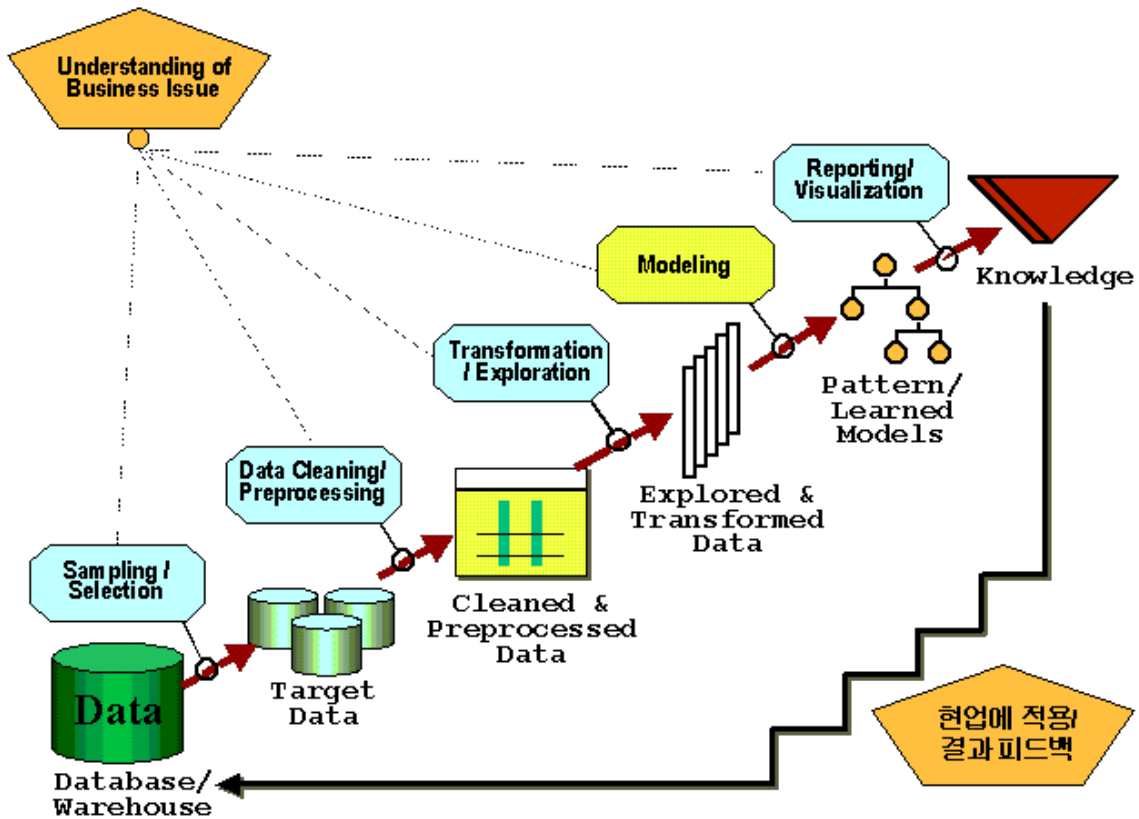
데이터 마이닝의 응용 분야로 신용평점 시스템(Credit Scoring System)의 신용평가모형 개발, 사기탐지시스템(Fraud Detection System), 장바구니 분석(Market Basket Analysis), 최적 포트폴리오 구축과 같이 다양한 산업 분야에서 광범위하게 사용되고 있다.

단점으로는, 자료에 의존하여 현상을 해석하고 개선하려고 하기 때문에 자료가 현실을 충분히 반영하지 못한 상태에서 정보를 추출한 모형을 개발할 경우 잘못된 모형을 구축하는 오류를 범할 수가 있다.

데이터마이닝은 데이터 분석을 통해 아래의 표와 같은 분야에 적용하여 결과를 도출할 수 있다.

분야	결과
분류(Classification):	일정한 집단에 대한 특정 정의를 통해 분류 및 구분을 추론한다.
군집화(Clustering)	구체적인 특성을 공유하는 군집을 찾는다. 군집화는 미리 정의된 특성에 대한 정보를 가지지 않는다는 점에서 분류와 다르다
연관성(Association):	동시에 발생한 사건간의 관계를 정의한다.
연속성(Sequencing):	특정 기간에 걸쳐 발생하는 관계를 규명한다. 기간의 특성을 제외하면 연관성 분석과 유사하다
예측(Forecasting)	대용량 데이터집합내의 패턴을 기반으로 미래를 예측한다.

이러한 데이터 마이닝은 아래 그림과 같이 반복된 연속 절차로 이루어진다.



(1) Sampling/Selection

방대한 양의 모집단인 데이터베이스에서 모집단을 닮은 작은 양의 데이터(Sample:표본)을 추출한다. 실제 존재하는 모든 데이터를 이용하여 어떠한 패턴을 찾아내는 것이 이상적이나 실제로 그렇게 하기엔 매우 큰 작업이 된다. 따라서 적당한 크기의 데이터를 이용하여 정확한 패턴을 찾아내는 기법이 필요하다.

(2) Cleaning/Preprocessing

다양한 데이터 소스들의 통합으로 데이터의 일관성을 유지하고 스키마를 통합 등 불완전한 데이터를 정제하여 데이터의 무결성과 질을 보장해야 한다. 실세계의 데이터베이스는 파일DB, DBMS 등에 따라 다양한 모습으로 존재하며 이러한 다양한 소스들의 데이터를 데이터 마이닝 엔진이 인식하기 쉬운 형태로 일관성 있게 통합이 필요하다. 그것이 데이터 스키마 및 포맷등을 통합하는 작업이다.

(3) Transformation/Exploration

분석에 필요한 데이터를 선별하고 수치화 작업을 통하여 수많은 변수들 간의 관계를 설정하는 단계이다. 데이터의 모양의 면밀히 검토하여 정보화의 기반을 다지는 과정으로 요약/집계 등의 연산을 통해 데이터 마이닝에 적합한 데이터로 변환한다.

(4) Data mining

데이터 마이닝에서 가장 중요한 단계로 데이터로부터 다양한 형태의 마이닝 기법을 적용시켜 패턴을 추출해 낸다. 마이닝의 결과를 얻기 위해 어떤 마이닝 기법을 구사할 것인지가 매우 중요하며 Neural Network, 결정 트리 등의 기법 사용된다.

(5) Reporting/Visualization

사용자들에게 데이터 마이닝의 결과를 그래프나 각종 차트등으로 사용자들에게 보기 편하고 이해하기 쉬운 형태로 제공하는 과정이다.

3. 협업적 필터링 기법

본 장에선 협업적 필터링 기법 중 사용자 기반 협업적 필터링 기법과 아이템 기반 협업적 기법의 진행 과정을 설명하고 협업적 필터링 기법이 일반적으로 가지는 문제점을 소개하며 그 중 희박성 문제점을 보완하기 위한 방법을 소개한다.

3.1. 사용자기반 협업적 필터링

사용자 기반 협업적 필터링 기법은 추천 대상자의 선호도 점수를 기반으로 추천 대상자와 비슷한 인접 이웃 그룹을 형성하고 그 인접 이웃 그룹으로부터 추천 목록을 생성하여 추천하는 방법이다. 사용자 기반 협업적 필터링 기법의 동작 과정은 다음과 같이 세 단계로 이뤄진다.

(1) 사용자-아이템 매트릭스 생성

추천 대상자가 선호도를 매긴 모든 아이템들에 대해서 다른 사용자들이 평가한 선호도 값으로 구성된 사용자-아이템 매트릭스를 생성한다. 만약 추천 대상자 user1이 item1, item3, item5, item8, item10에 대해서 선호도를 각각 (8, 10, 9, 7, 8)과 같이 매겼다면 아래와 같은 사용자-아이템 매트릭스가 생성된다.

	item1	item3	item5	item8	item10
user1	8	10	9	7	8
user2	4	-	3	5	8
user3	7	9	-	-	4
user4	9	10	-	7	6

(2) 사용자간 유사도 생성 및 인접 이웃 그룹 형성

앞 단계에서 생성한 사용자-아이템 매트릭스를 기반으로 추천 대상자와 다른 사용자들 간의 유사도를 측정한다. 유사도를 측정하는 데에 사용되는 공식으로는 피어슨 계수, 유클리드 거리, 자카드 거리, 코사인 계수 등을 이용하는 방법이 있다. Herlocker et al.(1999)은 고객들 간의 유사도를 계산할 때, Pearson Correlation Coefficient를 사용하는 것이 Cosine을 사용하는 것보다 높은 추천 성과를 보인다고 발표하였다. 따라서 본 논문에서는 유사도를 측정하기 위해 Pearson Correlation Coefficient를 사용하였다.

피어슨 계수를 이용한 유사도 값 $sim(A,B)$ 는 두 고객 A, B에 대해 사용자 A를 기준으로 측정한 사용자 B의 유사도 값이며 아래와 같은 공식으로 적용한다.

$$sim(a,b) = \frac{\sum_{i \in RI(a,b)} (r_{a,i} - \bar{r}_a)(r_{b,i} - \bar{r}_b)}{\sqrt{\sum_{i \in RI(a,b)} (r_{a,i} - \bar{r}_a)^2 \times \sum_{i \in RI(a,b)} (r_{b,i} - \bar{r}_b)^2}}$$

여기서 r_{ai} 는 사용자 A의 선호도 값을 말하며 $\overline{r_a}$ 는 사용자 A가 매긴 선호도 평균값이다.

목표고객에 대한 다른 고객들의 유사도를 계산한 후 목표고객에 대한 인접 이웃 그룹 형성은 다음의 두 가지 방법으로 형성할 수 있다. 첫 번째는 목표고객과의 유사도가 미리 설정한 임계값 이상인 고객들로 인접 이웃 그룹을 구성하는 것이고, 두 번째는 K-NN 기법으로 목표고객과의 유사도가 높은 상위 K명의 고객들로 인접 이웃 그룹을 형성하는 것이다.

(3) 평가치 예측 및 추천목록 생성

인접 이웃 그룹이 평가한 평가치를 이용하여 목표고객이 평가하지 않은 추천 후보 아이템들에 대해 평가치를 예측한 후 추천목록을 생성한다. 평가치 $R_{A,i}$ 는 고객 A의 아이템 I에 대한 평가치이며 목표고객 A에 대한 인접 이웃 그룹의 평가치들을 가중 평균하여 다음과 같이 예측한다.

$$R_{A,i} = \overline{R_A} + \frac{\sum_{j=1}^k \omega(A, j)(R_{j,i} - \overline{R_j})}{\sum_{j=1}^k |\omega(A, j)|}$$

여기서 $\overline{R_j}$ 는 고객 A에 대한 인접 이웃 그룹을 형성하고 있는 고객 j의 이용 가능한 평가치들의 평균값을 뜻한다. 목표고객이 평가하지 않은 아이템들의 평가치를 예측한 후 Top-N 기법으로 추천목록을 생성한다. 즉, 목표고객의 예측된 평가치 중에서 수치가 높은 상위 N개의 아이템을 목표고객에 대한 추천목록으로 생성한다.

3.2. 아이템기반 협업적 필터링

아이템기반 협업 필터링은 사용자 기반 협업 필터링과 매우 유사한 방법이다. 단지 비교 대상이 사용자가 아닌 아이템이 된다는 것이 차이점이다. 앞서의 사용자 기반 협업 필터링에선 유사도 측정을 통해 인접 이웃 그룹으로부터 평가치를 예측하여 추천 목록을 생성했지만 아이템 기반 협업적 필터링에선 아이템 간의 유사도를 계산하여 높은 유사도를 갖는 아이템을 추천 목록으로 생성한다.

아이템 기반 협업 필터링에서 유사도를 계산 할 아이템을 선정하는 방법이 중요한데 대표적인 두 가지 방법을 소개하면 첫 번째로 사용자가 과거에 구매한 아이템을 선택하는 방법이 있으며 두 번째로 사용자가 시스템 상에서 “상세보기”와 같은 행동을 통해 선택한 아이템을 유사도를 계산 할 아이템으로 선정하는 방법이 있다. 본 논문에선 사용자가 시스템 상에서 “상세보기” 메뉴를 통해 선택한 아이템을 유사도를 계산 할 기준 아이템으로 선정하여 추천 알고리즘을 진행한다.

3.3. 협업적 필터링의 문제점

협업적 필터링은 대부분의 추천 시스템에서 주요한 데이터 마이닝 기법으로 사용된다. 하지만 이러한 협업적 필터링은 희박성, 확장성 그리고 투명성의 문제점을 가지고 있다.

사용자기반 협업적 필터링에서 희박성은 콜드 스타트로 불리는 문제점으로 고객이 아이템의 부여한 선호도 정보가 부족하여 추천의 정확도가 떨어지는 문제점이다. 취급하는 상품의 개수가 많은 아마존이나 e-Bay와 같은 대영 쇼핑몰에서는 고객들이 실제로 경험하고 평가하는 아이템들의 개수가 전체의 1%도 안될 뿐 아니라 평가를 하지 않는 고객의 수도 많다. 즉, 협업적 필터링 과정 중 1단계에서 필요한 사용자-아이템 매트릭스의 대부분이 비어있는 현상이 발생한다. 그에 따라 정확한 유사도의 계산이 불가하여 추천 시스템의 성능이 저하된다.

아이템기반 협업적 필터링에서 희박성은 초기 평가자 문제로 불리는 문제점으로 아이템에 부여된 선호도의 정보가 충분하지 않아 추천의 정확도가 떨어지는 문제점이다. 선택된 아이템에 대해 다른 아이템들과 유사도를 계산할 때 사용자기반 협업적 필터링과 마찬가지로 사용자-아이템 매트릭스의 구성이 필요하다. 하지만 충분한 데이터가 존재하지 않아 대부분이 빈칸으로 존재하고 그로인해 추천 시스템의 성능이 저하된다.

확장성(Scalability)은 사용자의 수와 거래데이터의 개수가 늘어남에 따라 목표고객의 인접 이웃 그룹을 형성하기 위한 연산이 기하급수적으로 늘어난다는 문제점이다. 예를 들어, 협업 필터링이 만 명의 목표고객들에게 추천을 하고자 한다면, 만 명의 목표고객들의 인접 이웃 그룹을 형성하기 위해 수십 만 명의 고객데이터와 이들이 부여한 수 백만 개의 선호도 정보를 검색해야 한다. 추천시스템이 상대하는 고객의 수나 취급하는 아이템의 개수가 늘어나면 인접 이웃 그룹을 형성하기 위한 계산의 양이 기하급수적으로 늘어나서 추천목록을 생성하기까지 오랜 시간이 걸려 추천 시스템의 효율성이 떨어지게 된다. 확장성의 문제를 완화하기 위해 개별 고객들 간의 유사도를 계산하는 대신, 아이템들 간의 유사도를 계산하거나 고

객을 군집화한 후 유사도를 계산하는 방법들이 제안되었다.

투명성(Transparency)은 추천 결과와 고객의 선호도와의 관계가 불명확하다는 문제점이다. 협업필터링은 수식계산을 포함하고 있는 블랙박스를 통해 고객에게 맞는 추천목록을 생성하고 제시하기 때문에, 고객은 아이템들이 자신에게 어떠한 방법을 통해 추천 되었는지를 이해하는 것이 어려울 수 있다.

3.4. 협업적 필터링의 문제점 보완

본 논문에선 협업적 필터링에서 갖는 희박성 문제점의 해결에 대해 개인의 특성을 이용한 인구 통계학적 추천방법을 통해 해결한다. 인구 통계학적 추천방법은 다양한 종류의 항목이나 범주에 대한 사용자의 성향을 분석하기 좋은 방법으로 구매 이력과 같은 데이터가 없는 초기단계에서 갖는 협업적 필터링의 희박성 문제를 효과적으로 해결할 수 있는 방법이다. 재능 추천 시스템에서 사용되는 개인의 특성은 직종, 연령대, 성별, 전공, 혈액형, 거주지 총 6가지의 특성을 이용하며 각각의 특성에 (20%, 15%, 10%, 20%, 15%, 20%)의 가중치를 부여하여 사용자간의 유사도를 측정하며 각각의 특성이 갖는 범주는 아래와 같다.

20%	15%	10%	20%	15%	20%	
직종	연령대	성별	전공(희망전공)	혈액형	거주지	
요인	요인	요인	요인	요인	요인	
초중고	10대미만	남	인문학	A	서울	
대학생	10대	여	경영학	B	경기	
대학원생	20대		공학	O	인천	
회사원	30대		자연대	AB	강원	
디자이너	40대		의과학		충남	
프로그래머	50대		교직과정		대전	
방송/광고	60대이상		신학		충북	
자영업			체육학		전북	
공무원					전남	
정치인					광주	
교직					경남	
교수					경북	
의료인		부산				
군인		대구				
운동선수		울산				
종교인		제주도				
연예인						
주부						
농수축산업						
기타						

4. 개인의 특성을 이용한 협업적 필터링 기반의

개인화 재능 추천 시스템의 설계

본 장에서는 협업적 필터링과 이를 보완하기 위한 인구 통계학적 기반 필터링이 적용된 개인의 특성을 이용한 협업적 필터링 기반의 재능 추천 시스템에 대해서 설계하며 설계를 위해 전체 시스템의 구조와 재능 추천 시스템의 구성도를 보인다.

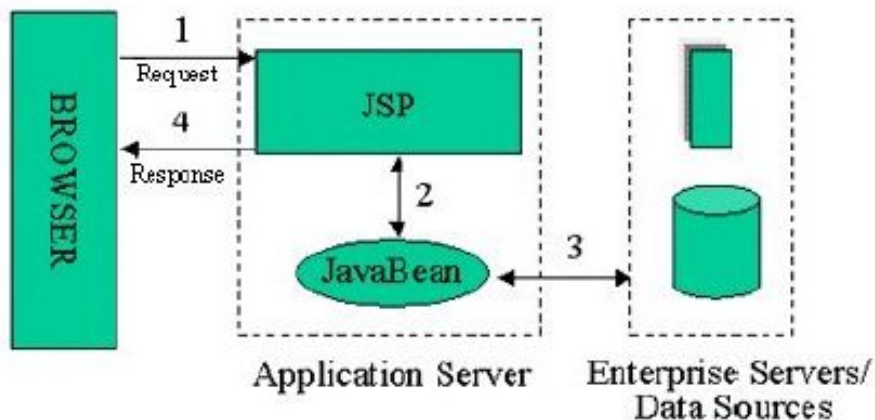
4.1. Model2 기반의 JSP환경의 서버-클라이언트 구조

JSP는 유닉스나 리눅스 환경에서 주로 사용되는 웹 서버 언어로 웹 서버 프로그램은 아파치와 웹 어플리케이션 서버인 Tomcat과 데이터베이스인 MySQL과 연동에 동작한다. JSP는 정적인 HTML 문서의 기능을 보완하는 역할을 하는 것으로 HTML 문서 내에 JSP코드를 추가하거나 jstl, el과 같은 별도의 태그 형식으로 사용한다.

클라이언트가 웹 페이지를 요청하면 서버는 해당 JSP 페이지를 서블릿의 형태로 컴파일하며 이때 일어나는 컴파일은 최초한 번만 수행된다. 그 후 해당 서블릿이 반환하는 HTML 형식으로 이루어진 웹 페이지를 반환하여 클라이언트에 전달한다.

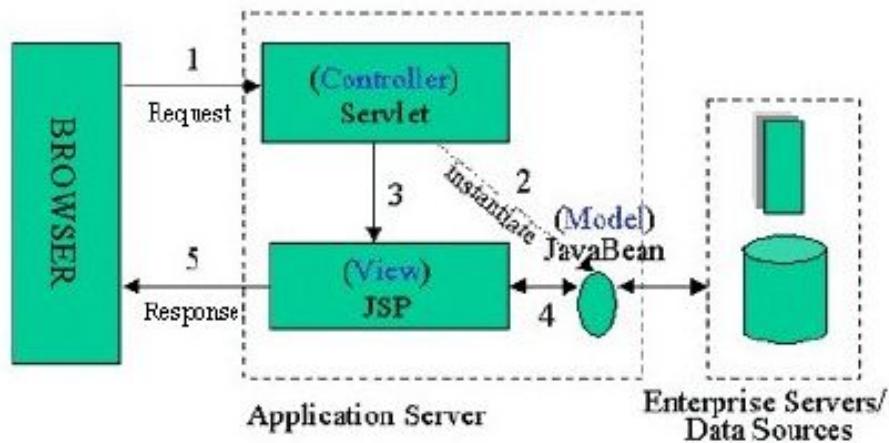
JSP와 비슷한 기능을 수행하는 ASP(Active Server Page) 또는 ASP.NET은 MS 윈도우즈 서버 운영체제에서 제공하는 다양한 컴포넌트를 활용할 수 있다는 장점이 있다. 대부분의 일반 사용자는 윈도우즈 운영체제에 익숙하기 때문에 사용하기 편리하지만, 유닉스나 리눅스 등의 운영체제에서는 사용할 수 없다는 단점이 있다. 그에 비해 PHP는 모든 종류의 운영체제에서 지원하며, 공개용 데이터베이스인 MySQL과도 연동이 쉽다.

JSP를 이용한 개발 방식에서 Model1 구조를 이용하는 방법과 Model 구조를 이용하는 방법이 있다. Model1 개발 방식은 JSP/Servlet이나 JSP와 java Bean을 이용하여 개발하는 방식이며 MVC패턴에 적용시키면 Model은 데이터베이스가 담당하고 View와 Controller의 역할은 JSP가 담당하는데 Controller의 의미는 매우 미미하다. Model1 개발 방식의 동작은 아래와 같다.

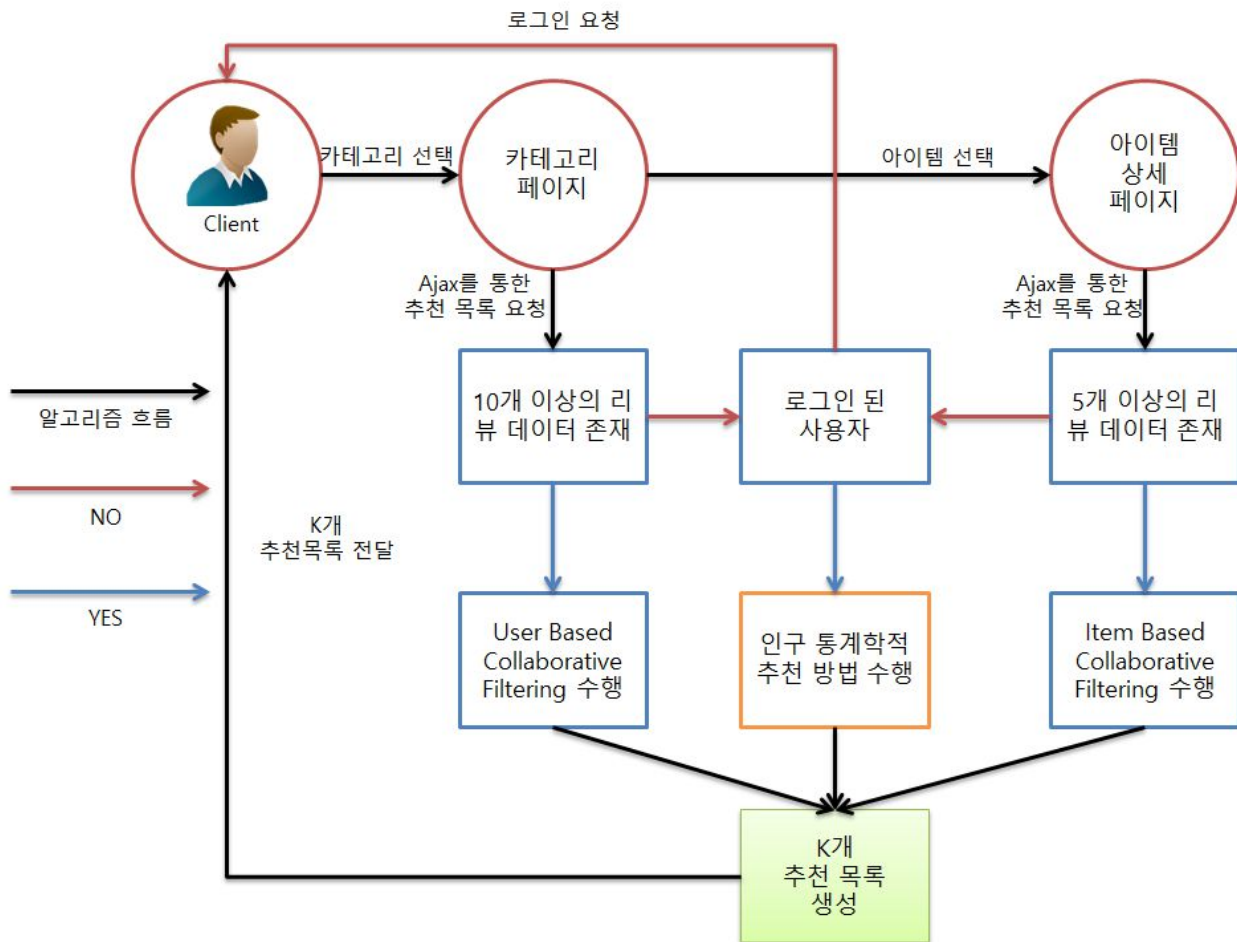


Model2 개발 방식은 모바일 등과 같은 다양한 애플리케이션의 개발 방식에 있었던 개념을 웹에 적용한 개념으로 MVC패턴에서 각각의 역할을 확실하게 나누어 작업하고자 하는

일을 분담시키기 위한 결과물이다. Model의 역할은 애플리케이션의 로직을 담당하는 부분으로 데이터베이스나 Legacy System과의 로직을 담당하는 부분이다. 주로 DTO(Data Transfer Object), DAO(Data Access Object)의 자바빈즈가 이 역할을 한다. View는 사용자가 직접 사용하는 부분을 말하며 Presentation 로직을 담당하는 부분이다. Controller와 Model에 의해 생성된 결과물을 보여주며 주로 JSP페이지가 이 역할을 한다. Controller는 클라이언트의 요청의 진입점으로 서비스를 호출하며 요청 파라미터를 저장하는 역할로 사용자의 요청을 바당 요청에 해당하는 작업 후 작업 결과에 따른 응답을 하는 역할을 한다. 즉, Model과 View사이에서 양쪽과의 커뮤니케이션을 담당하며 주로 Servlet이 이 역할을 담당한다. Model2 개발 방식의 동작은 아래와 같다.



4.2. 개인화 재능 추천 시스템의 구성도



시스템에서 사용되는 추천 알고리즘 방식은 최초 데이터가 부족한 경우 사용되는 인구 통계학적 필터링 기법과 사용자 기반 협업적 필터링, 아이템 기반 협업적 필터링 세 가지로 구성되어 있다. 협업 필터링 기반의 개인화 재능 추천 시스템에서는 사이트에 가입한 고객들을 대상으로 서비스가 이루어지며 고객들은 최초 가입 시 자신의 나이, 성별, 직종, 전공, 혈액형, 거주지 정보를 입력받아 인구 통계학적 필터링 기법에 사용 할 수 있도록 데이터를 생성한다. 이것은 협업적 필터링에서 한계점인 희박성 문제를 해결하기 위한 것으로 초기 평점 데이터가 적을 경우 추천에 대한 정확도나 만족도가 떨어지는 것을 방지하기 위한 것이다.

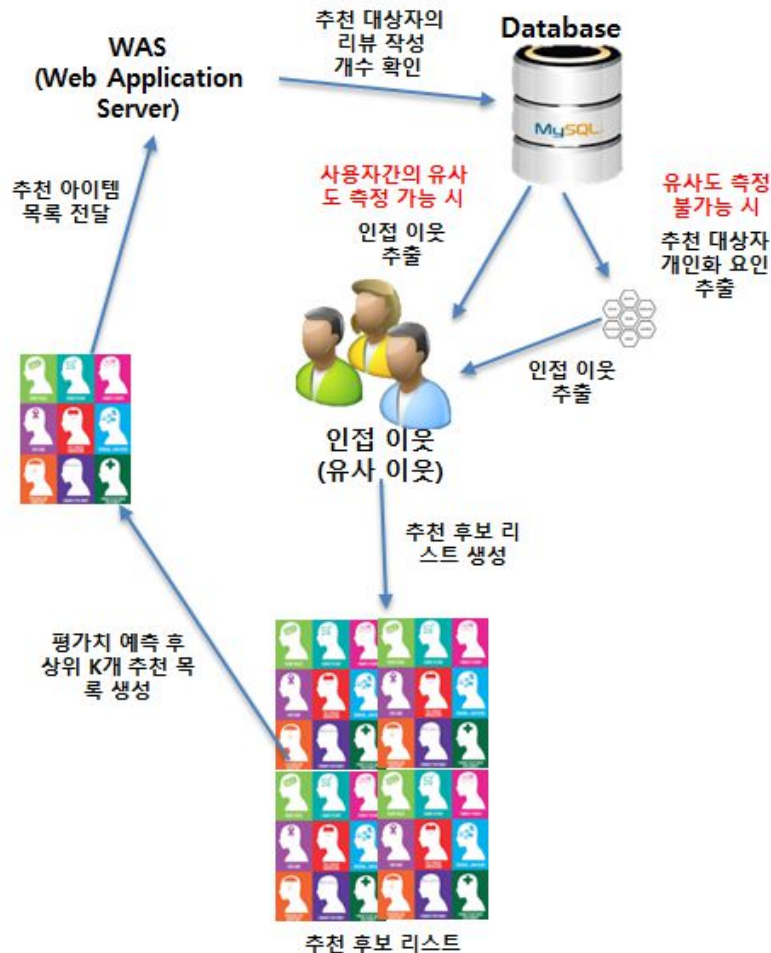
초기 평점에 대한 데이터가 부족한 경우의 사용자 기반 협업적 필터링에선 인구통계학적 필터링을 이용하여 인접 이웃의 그룹을 형성하는 것을 통해 희박성을 해결한다. 또한 초기 평점에 대한 데이터가 부족한 경우의 아이템 기반 협업적 필터링에선 마찬가지로 인구통계학적 필터링을 이용한 방법으로 추천 알고리즘이 대체되며 사용자 기반 협업적 필터링에서와 같이 인접 이웃 그룹의 형성을 통한 희박성 해결로 이루어진다.

서비스를 이용하면서 고객은 자신이 구매 행위를 한 재능에 대해 평점 정보를 등록하게 되며 그것이 희박성 문제를 해결 할 수 있는 최소값에 도달하면 시스템은 자동으로 협업적 필터링 기법으로 추천 서비스를 변경하고 사용자에게 추천 목록을 반환한다.

4.3. 인접 이웃 그룹 형성 및 추천 목록 생성 과정

(1). 사용자 기반 협업적 필터링 시나리오

사용자 기반 협업적 필터링 시나리오

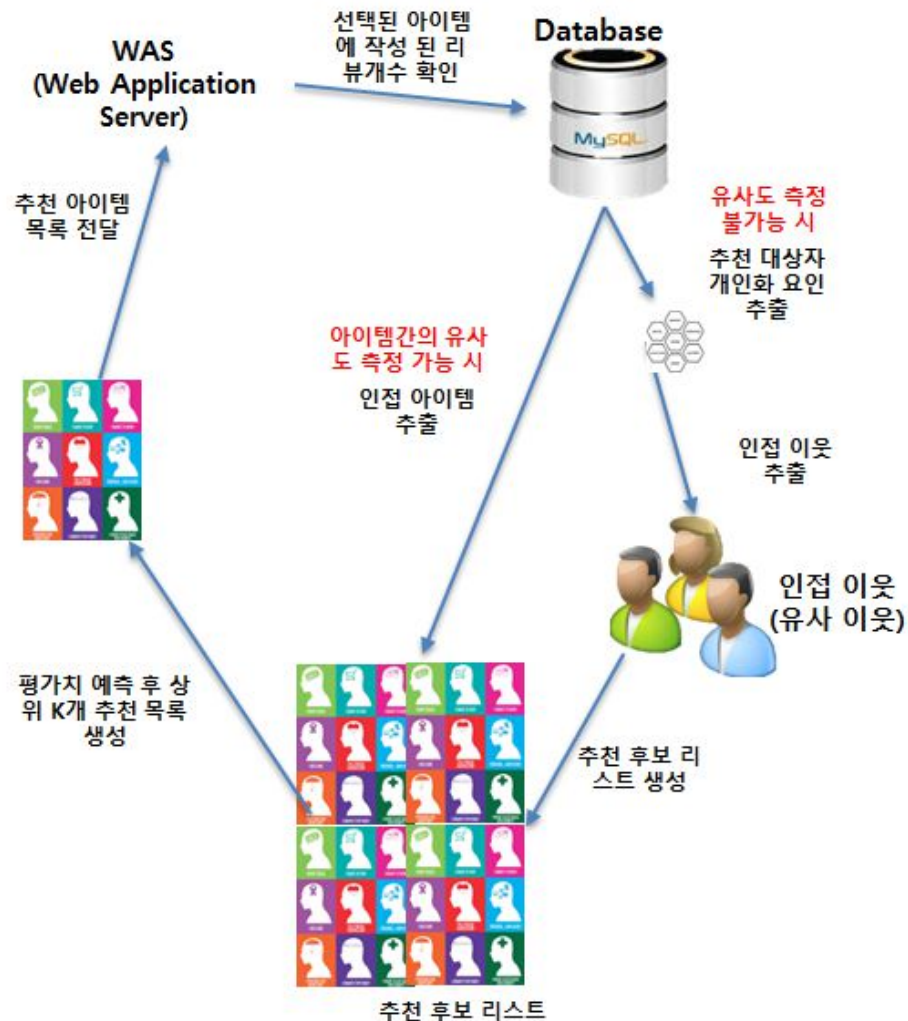


사용자가 최초 가입한 경우 사용자의 평점 데이터는 존재하지 않으며 평점 데이터는 서비스 이용 중 사용자의 행동에 따른 입력을 통해 축적된다. 이때는 사용자 기반 협업적 필터링의 기반이 되는 평점 데이터가 부족하기 때문에 희박성 문제가 발생하여 적절한 인접 이웃 그룹의 형성이 불가하고 그로 인해 추천 결과의 신뢰성의 매우 낮아진다. 이것을 해결하기 위해 회원 가입 시 입력한 나이, 성별, 전공, 직종, 혈액형, 거주지를 기반으로 각각의 가중치를 두고 인구통계학적 필터링 기법을 적용한다. 성향이 유사한 사용자들을 그룹화 하여 인접 이웃 그룹으로 형성하고 그룹군의 평점 정보를 통해 추천 목록을 생성한다.

여기서 인구통계학적 기법에서 협업적 필터링 기법으로의 전환, 즉 협업적 필터링을 위한 평점 데이터의 최소치는 사용자의 평점 데이터 10개로 정한다. 사용자가 입력한 평점 데이터가 10개 이상인 경우 그 평점 정보를 기반으로 사용자 기반 협업적 필터링이 진행되며 10개 미만인 경우 사용자가 회원 가입 시 입력했던 개인의 특성 데이터를 기반으로 인구 통계학적 기반 필터링을 통해 추천 목록을 생성한다.

(2). 아이템 기반 협업적 필터링 시나리오

아이템 기반 협업적 필터링 시나리오



최초 등록된 아이템의 경우 아이템에 등록된 평점 데이터는 존재하지 않는다. 아이템의 평점 데이터는 서비스 가동 중 사용자의 행동에 따라 축적될 것이다. 아이템간의 유사도를 측정하기 위한 최소의 평점 데이터가 없다면 아이템 기반 협업적 필터링에서 유사한 아이템을 찾아낸 결과가 매우 부정확해지고 추천 결과의 신뢰성이 떨어진다.

이것을 보완하기 위해 앞서의 사용자 기반 협업적 필터링과 같이 사용자의 개인 특성을 이용하여 인구통계학적 필터링 기법으로 추천 기법을 대체한다.

인구통계학적 기법에서 협업적 필터링 기법으로의 전환, 즉 협업적 필터링을 위한 평점 데이터의 최소치는 아이템에 등록된 평점 데이터 5개로 정한다. 사용자가 해당 아이템에 등록한 평점 데이터가 5개 이상인 경우 아이템 기반 협업적 필터링이 진행되며 그렇지 않은 경우 인구통계학적 필터링을 통해 추천 목록을 생성한다.

4.4. 희박성 문제의 해결

신규가입을 한 사용어나 새로 등록된 재능의 경우 평점에 대한 데이터가 존재하지 않아 다음과 같이 유사도를 이용한 그룹화를 할 수 없는 희박성의 문제가 발생한다.

재능 상세정보	A/S 및 환불/배송 규정	재능 리뷰
---------	----------------	-------

사용자 리뷰

등록된 리뷰가 없습니다.

당신에게 적합한 맞춤 재능들 입니다.

이 추천이 유

평점 데이터 부족으로 인해 추천을 진행할 수 없습니다.

로그인하면 멋진 추천 재능목록을 확인할 수 있습니다.

따라서 다음과 같이 회원가입 시 입력받는 나이, 성별, 혈액형, 거주지, 직종, 전공의 항목들에 대해 각 항목별 가중치를 달리 두어 해당 사용자와 유사한 사용자들을 인접 이웃 그룹으로 형성하고 그 인접 이웃들로부터 추천 목록을 생성해내어 희박성의 문제를 해결한다.

®Recommend Your Talent 아이디 비밀번호 [로그인](#) [회원가입](#) [주요기능](#)

시스템 소개 재능관리 구매목록 커뮤니티 구매현황 공지사항 ! Help 카테고리

Talent Recommend System

기타

프로필 사진 [찾아보기...](#)

※정확한 추천을 위해 필요합니다.

직종

나이 세

성별 ☐ 남성 ☐ 여성

전공(희망전공)

혈액형

[확인](#) [뒤로](#)

5. 개인의 특성을 이용한 협업적 필터링 기반의

개인화 재능 추천 시스템의 구현

본 장에서는 협업적 필터링을 이용한 개인화 재능 추천 시스템의 구현 환경에 대해 기술하고 협업적 필터링을 이용한 개인화 재능 추천 시스템의 실제 구현된 웹 페이지와 그에 대한 설명을 기술한다.

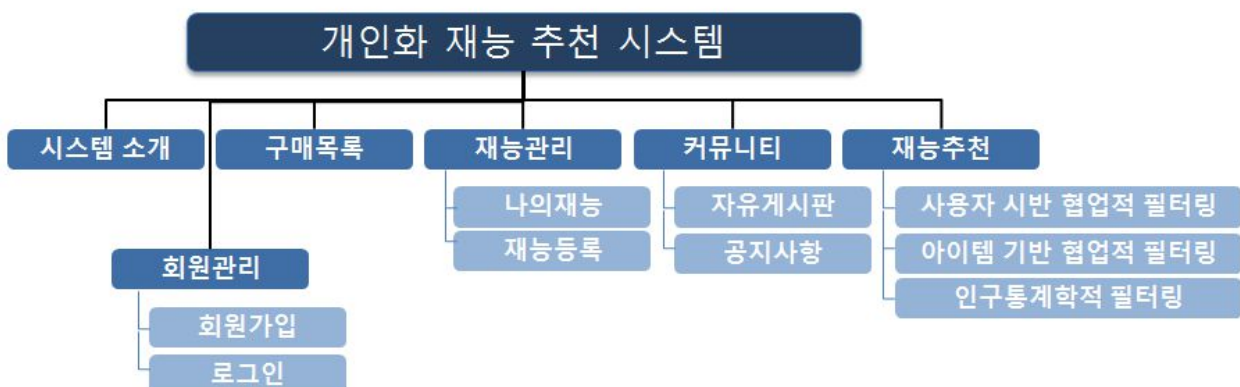
5.1. 구현 환경

아래의 표는 전체 시스템의 구현 환경을 나타낸 것으로 서버의 운영체제는 Linux Ubuntu12.04 LTS를 사용했고 웹 서버는 Apache 2.2, WAS는 tomcat8을 사용했다. 데이터를 저장하기 위한 Database는 MySQL Server 5.5를 사용했다.

시스템 구성 요소	시스템 세부 내용
Server OS	Linux Ubuntu12.04 LTS
Database	MySQL 5.5
Web Server	Apache 2.2
Web Application Server	Tomcat 8
Web Language	JSP & Servlet
jQuery	jQuery-2.1.4
Java	Java 8
JDK	jdk 1.8.0_92

5.2. 개인화 재능 추천 시스템 구현

전체적인 개인화 재능 추천 시스템의 웹 페이지 구성도는 아래와 같다.



다음 그림은 공지사항 페이지이다. 공지사항 페이지에는 관리자만이 공지할 사항을 게시할 수 있으며 상위에 가장 최신의 글이 노출된다.

공지사항

번호	<input type="checkbox"/>	제목	글쓴이	조회
공지	<input type="checkbox"/>	홈페이지 이용규칙 안내.	관리자	37
1	<input type="checkbox"/>	홈페이지 이용규칙 안내.	관리자	37

순천향대학교 (SoonChunHyang University)

컴퓨터 소프트웨어공학과 (Department of Computer Software Engineering)

20114016 이 태수

dlxotn12345@naver.com



순천향대학교



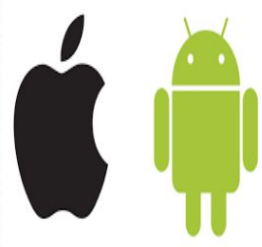





Talent Recommendation System

나의 재능 페이지는 아래와 같다. 내가 등록한 재능 목록을 볼 수 있으며 수정 및 삭제의 재능 관리가 가능하다.

나의 재능

(클릭하면 상세 내역을 볼 수 있습니다.)

최근등록순 ▲ 등록순 ▼

<p>재능 등록일 : 2016-01-12 12:00 가격 : ₩ 13,500</p>  <p>[IT&컴퓨터 > 어플리케이션] 어플리케이션 개발 작업 도와드립니다. 다양한 어플리케이션 개발 경험</p> <p>  </p> <p>재능 수정 재능 삭제</p>	<p>재능 등록일 : 2016-01-12 12:00 가격 : ₩ 130,000</p>  <p>[IT&컴퓨터 > 어플리케이션] iOS기반 게임 제작해드립니다. 게임 개발 경험은 적지만 최선을 노력으로</p> <p>  </p> <p>재능 수정 재능 삭제</p>	<p>재능 등록일 : 2016-01-12 12:00 가격 : ₩ 1,300,012</p>  <p>[IT&컴퓨터 > 웹] JSP 기반 간단 개인 웹페이지 제작 유지보수 불가</p> <p>  </p> <p>재능 수정 재능 삭제</p>
--	--	---

다음 그림은 자유게시판 화면이다. 공지사항 게시판과 다른 UI로 제작되어있으며 300글자 이내의 미리보기가 지원된다.

@recommend Your Talent
admin ▼

시스템 소개재능관리구매목록커뮤니티구매현황공지사항! Help

전체 게시물 : 39 | 전체 페이지 : 8
카테고리
글쓰기메인으로

43번째 글입니다.

admin
2016-01-12

개인화요소는 [0]

나이 10대 20대 30대 40대 50대
성별 남 여
전공 1 공학전공, 인문학전공, 자연계전공, 사회계열...
전공 2
성향 소독
커뮤니티 활동수준 하 중하 중 중상 상
자세히 보기...

2번 조회됨

다음 그림은 카테고리 상세 페이지이다. 해당 카테고리에 포함되는 아이템 목록을 보여주며 이름순, 등록순의 정렬 보기가 가능하다.

@recommend Your Talent
admin ▼

시스템 소개재능관리구매목록커뮤니티구매현황공지사항! Help

카테고리

나의 재능으로 수익을 창출해보세요.

IT&컴퓨터
재능 도움요청
이름순 ▼
등록순 ▼

IT&컴퓨터 > 웹

웹 (6)
어플리케이션 (4)
워드프레스 (6)
퍼블리싱 (2)
일반소프트웨어 (4)
커머스 & 쇼핑몰 (5)
게임제작 (0)
임베디드 (0)
기타 (2)

구분
☐ 일반사이트
☐ 웹 프로그래밍
☐ 데이터베이스
☐ SNS
☐ 블로그

작업유형
☐ 웹사이트 제작






₩ 1,300,012
JSP
Java Server Pages
JSP 기반 간단 개인 웹페이지 제작 (7) 15,11,10,7,2,1,
유지보수 불가
admin ★★★★★ 8.0점

₩ 80,000
python
파이썬 기반 웹페이지 설계/구현 (11) 19,14,7,2,
데이터베이스 설계 가능
admin ★★★★★ 9.0점

₩ 21,000
Blog
블로그 운영/관리 해드립니다. (12) 23,20,16,10,9,5,
블로그 전문
admin ★★★★★ 8.0점

다음은 주문내역 및 구매목록 화면이다. 사용자가 구매하거나 주문한 내역에 대한 조회가 가능하다.

- 20 -

®ecommend Your Talent						test10 ▼
시스템 소개	재능관리	구매목록	커뮤니티	구매현황	공지사항	! Help
주문내역 및 구매목록						카테고리
번호	주문일	상품정보	수량	금액(W)	추천여부	
39	2016-04-17 15:42	 iOS기반 게임 제작해드립니다.(4)	1	₩ 13,221	X	
38	2016-04-17 15:42	 HTML5] 웹페이지 UI제작(10)	1	₩ 15,000	X	
37	2016-04-17 15:27	 어플리케이션 개발 작업 도와드립니다.(2)	1	₩ 13,500	X	
36	2016-04-17 15:26	 JSP 기반 간단 개인 웹페이지 제작(7)	1	₩ 1,300,012	X	
35	2016-04-17 15:22	 표현의 새로운 시작, 로고송 & CM송 (사가,프로포즈곡,자신만의 노래) 제작해 드립니다 (1)	2	₩ 26,000	X	

다음 그림은 재능 상세보기 페이지이다. 사용자가 등록한 재능 정보와 판매자 정보 등이 제공된다.

®ecommend Your Talent

test10 ▼

시스템 소개

재능관리

구매목록

커뮤니티

구매현황


공지사항

! Help

카테고리

"CF 애니메이션 제작해드립니다."(92)

CF 애니메이션 제작해드립니다.



재능마켓

재능관리

재능관리

재능관리

CF 애니메이션 제작해드립니다.

₩ 300,000

최소 소요기간 10 일

최대 주문개수 2 개

namiw

판매자 정보 보기

작업 수 2 개

주문 수 1 회

평점 0 점

리뷰 수 1 개

판매자 프로필

✓ 재능 주문

뒤로가기

다음 그림은 추천시스템의 추천 성능 조회 페이지이다. 일정 기간 동안의 구매 이력 중 각각의 추천 알고리즘에 의해 유도된 구매 횟수와 추천 만족도 조사 내역을 기반으로 성능을

보여준다.

Performance of Talent Recommendation System

'협업 필터링 기반 재능추천 시스템의 성능 현황입니다.'

추천 성능

Performance of Talent Recommendation System

20114016
이태수

기간 선택

2016년 1월 1일

2016년 6월 9일

조회

성능조회를 위한 기간을 좌측 아래 메뉴에서 선택해주세요.

총 59회의 사용자 만족도 평가조사 중
12회의 사용자 만족 평가를받아
20%의 사용자 만족도가 측정됩니다.

총 121번의 구매 중
22번의 구매가 추천시스템에 의해 유도된 구매이며
전체 구매 활동 중20%의 비율을 차지합니다.

알고리즘 명	유도 구매 횟수
Personalization	0회
ItemBasedCF	14회
UserBasedCF	8회