

2014년도
학사학위논문

협업 필터링과 개인 성향을 이용한
차량 추천 시스템

Car Recommendation System Based on
the Collaborative Filtering and Personal
Propensity

2014년 10월 28일

순천향대학교 공과대학
컴퓨터소프트웨어공학과

모 하 영

2014년도
학사학위논문

협업 필터링과 개인 성향을 이용한
차량 추천 시스템

Car Recommendation System Based on
the Collaborative Filtering and Personal
Propensity

2014년 10월 28일

순천향대학교 공과대학
컴퓨터소프트웨어공학과

모하영

협업 필터링과 개인 성향을 이용한
차량 추천 시스템

Car Recommendation System Based on
the Collaborative Filtering and Personal
Propensity

지도교수 박 두 순

이 논문을 공학사학위 논문으로 제출함

2014년 10월 28일

순천향대학교 공과대학
컴퓨터소프트웨어공학과

모하영

모 하 영의 공학사학위논문을 인준함

2014년 10월 28일

심 사 위 원 인

순천향대학교 공과대학

컴퓨터소프트웨어공학과

차 례

제 1 장 서론	1
제 2 장 관련 연구	3
2.1 데이터마이닝(Data Mining)	3
2.2 개인화 추천 시스템	4
2.2.1 사용자 기반 필터링(User-based filtering)	5
2.2.2 내용 기반 필터링(Content-based filtering)	5
2.2.3 인구통계학적기반 필터링(Demographic-based filtering)	5
2.2.4 협업 필터링(Collaborative filtering)	6
제 3 장 협업필터링과 개인성향을 이용한 차량 추천 시스템 설계	10
3.1 차량 추천 시스템 구성도	10
3.2 희박성 문제 해결 및 추이 과정	11
제 4 장 협업필터링과 개인성향을 이용한 차량 추천 시스템 구현	13
4.1 구현 환경	13
4.2 시스템 구현 화면	14
제 5 장 결론 및 향후 연구과제	18
참 고 문 헌	19
부 록	20

그림 차례

<그림 2-1 데이터마이닝 과정>	3
<그림 2-2 데이터 마이닝의작업 흐름도>	5
<그림 3-1 시스템 구조도>	10
<그림 3-2 차량 추천 방법>	10
<그림 3-3 유사 사용자 선정>	11
<그림 3-4 유사그룹 형성 과정>	12
<그림 3-5 회박성 문제 해결과정>	12
<그림 4-1 차량 추천 시스템 구성도>	15
<그림 4-2 기아차량 정보화면>	16
<그림 4-3 현대차량 정보화면>	16
<그림 4-4 삼성 차량 정보화면>	17
<그림 4-5 회원가입시 입력할 정보>	17
<그림 4-6 이용자에 따라 다른 차량이 추천>	18

표 차 례

[표 2-1]협업필터링에서 고객 x 항목 매트릭스	6
[표3-1 각 카테고리별 가중치]	12
[표 4-1 시스템 구현 환경]	13

국문 초록

2014년 현재 국민들의 소득이 증가하며, 인터넷 보급률이 올라감에 따라 사람들이 원하는 정보를 얻기 위해 소요되는 시간과 노력을 절약하기 위한 방안으로 추천 시스템의 필요성이 강조되고 있다. 이로 인해 기업들은 고객이 인터넷을 통한 활동을 하면서 만들어진 데이터를 이용해서 얻은 수많은 데이터 중 유용한 정보를 얻기 위해 데이터마이닝을 추천시스템에 적용하게 되었다.

추천 시스템은 데이터마이닝을 통해서 이뤄지는데, 이중 가장 일반적으로 많이 사용되어지는 기법은 협업 필터링 기법이다. 협업 필터링 기법이라 함은 추천시스템 분야에서 가장 성공적으로 사용되어지는 기법으로 전자상거래 포털에서 가장 널리 이용되어지고 있는 기법이다. 하지만 협업 필터링 기법을 사용할 때 몇 가지 문제점이 발생할 수 있다. 그 중에서 하나가 희박성(Sparsity)이다.

희박성이란 새로운 이용자가 추천을 받으려 할 때, 협업필터링 기법만으로는 유사한 회원을 찾기 어렵고, 만약 찾아도 정확도가 떨어진다. 이를 해결하기 위해 개인 성향과 협업필터링 기법을 동시에 사용하여 희박성 문제를 해결한다.

본 논문에서 사용한 개인 성향으로는 탑승 수, 연봉, 성별, 직급을 이용했다. 위에 있는 개인 성향을 이용한 이유는 탑승 수에 따라 추천 추천되는 차량의 종류가 달라져야 하며, 연봉이나 성별에 따른 선호하는 차량이 달라진다고 생각하기 때문이다.

본 논문에서는 협업필터링의 희박성 문제를 해소하기 위하여 개인성향을 함께 이용하여, 추천 시스템의 성능을 향상 시켜 사용자에게 차량을 추천해주는 시스템을 설계하고 구현한다.

주요어 : Data Mining, 협업필터링, 추천 시스템, 개인성향

제 1장 서론

정보화시대가 도래하면서 전자상거래 시장의 변화 또한 빠른 속도로 변화를 맞이했다. 기존 오프라인의 고객층이 시간과 장소에 제약이 없는 온라인으로 이동하면서 더욱 많은 소비를 이끌어 낼 수 있게 되었지만, 수많은 정보 안에서 고객이 진정으로 원하는 정보를 얻기 위한 시간과 노력이 오히려 많이 필요하게 되었다. 이에 따라 이러한 시간과 노력을 절약할 수 있는 시스템의 필요성이 대두되어 지고 있다.

이러한 추천시스템을 만들 때 가장 많이 사용되는 기법으로는 협업필터링 기법이 있다. 협업필터링 기법은 지금까지 사용자의 높은 만족도를 이끌어 낼 수 있는 기법이었으나 몇 가지 한계점을 가지고 있다. 이는 희박성, 확정성 등의 문제점이다. 그 중에서 희박성이란 신규 회원이 차량을 추천 받으려 할 때, 협업필터링기법만을 사용할 경우 유사한 회원을 찾기 어렵고, 만약 찾아도 정확도가 떨어지기 때문에 이를 해결하기 위해 개인 성향을 추가로 이용하였다. 경우 기존에 있던 때 사용자가 가지는 만족도와 신뢰도를 올릴 수 있다. 이렇게 개인성향을 추가로 이용하여 사용자에게 차량을 추천해 주는 이유는 현재로써 추천 시스템에 가장 많이 사용되는 협업 필터링과 그 협업 필터링을 사용하면 생기는 희박성 문제를 개인 성향을 이용한 추천 시스템과 동시에 사용하여 해결하고자 함이다.

이 논문에서는 협업필터링 기법을 이용하여 차량을 추천하는 시스템을 만들고자 하는데 앞에 언급한 협업 필터링 기법의 희박성을 보완하기 위해서 내용 기반 필터링 기법을 함께 이용하여 사용자에게 알맞은 차량을 추천하는 시스템을 설계한다.

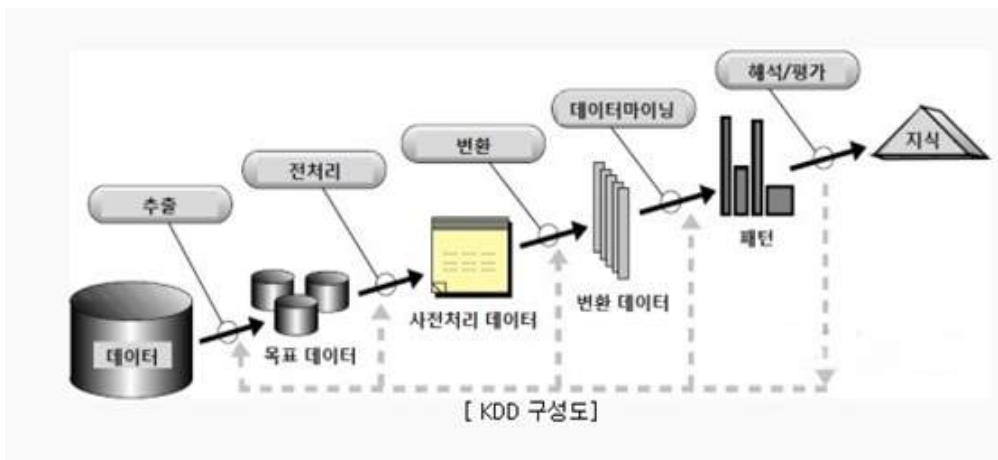
제 2장 연구 배경 및 관련 사례

본 장에서는 협업필터링을 이용한 차량 추천 시스템을 설계하고 구현하는데 필요한 데이터마이닝(Data Mining), 개인화 추천 시스템(Personal Recommendation System)에 대해 설명한다.

2.1 데이터마이닝(Data Mining)

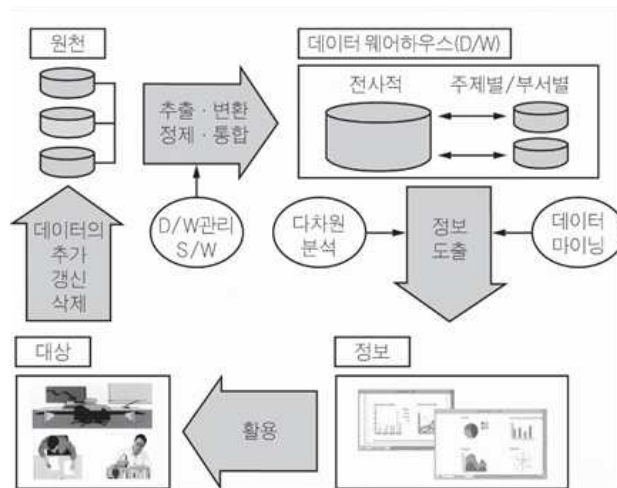
데이터마이닝(Data Mining)이란 대량의 데이터 가운데에서 유용한 정보를 추출하는 것으로 의미 있는 패턴과 규칙을 찾는 것이다. 다른 말로는 KDD(Knowledge Discovery and Data Mining)이라고도 일컫는다. 데이터마이닝에 대한 정의는 다양하나 대부분의 의미는 매우 큰 데이터베이스로부터 유용한 관계를 발견 또는 찾는다는 의미를 가지고 있다[1,2,3].

데이터마이닝은 일반적으로 <그림 2-1>과 같이 데이터 선택, 데이터 정제, 데이터 변환, 데이터마이닝, 패턴 평가, 지식 표현의 6단계로 실행된다[4].



<그림 2-1 데이터마이닝의 과정>

<그림 2-2>는 데이터마이닝의 작업 흐름도를 보여준다, 이는 데이터 선택, 데이터 정제, 데이터 변환, 데이터 추출, 데이터 통합등의 과정은 시간을 많이 요구하는 작업이다. 데이터 패턴 추출을 위해 데이터마이닝 알고리즘이 적용되는 단계는 데이터마이닝 전 과정 중 핵심 단계로 일반적으로 데이터마이닝을 하는 작업의 유형에 따라 연관규칙, 군집화, 사례기반 추론 등의 알고리즘이 사용될 수 있으며 복수의 기법들이 사용되기도 한다[5].



<그림 2-2 데이터마이닝의 작업 흐름도>

2.2 개인화 추천 시스템(personalized recommendation system)

개인화(Personalization)는 상인이 소비자의 이름, 관심사, 과거 구매력을 기반으로 시장에 전달할 메시지를 조정하여 특정 고객에 맞는 마케팅 메시지를 만들어내는 것을 말한다[6].

개인화 기법은 추천 시스템의 한 방법으로 사용자로부터 입력받은 명시적/묵시적 데이터를 바탕으로 개개인의 성향을 반영하여 추천의 질을 높이는 기술을 의미한다. 대표적인 개인화 기법은 다음과 같다.

2.2.1 사용자 기반 필터링(user-based filtering)

사용자 기반 필터링(user-based filtering)은 웹을 기반으로 하는 인터넷 쇼핑몰에서 이용되고 있는 성공적인 상품추천기법 중의 하나로서, 목표 고객과 유사한 구매이력을 보이는 이웃 고객들의 상품에 대한 선호를 바탕으로 목표 고객에게 유용한 상품을 추천하는 방법이며, 고객 선호도데이터가 이진 데이터인 경우, 고객간의 유사도는 다음의 식(1)을 통해서 구한다[7][8].

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad \text{식(1)}$$

2.2.2 내용 기반 필터링(content-based filtering)

사용자의 선호 정보를 제공받아 그와 유사한 항목을 추천해주는 기법[9]이다. 이는 사용자의 프로파일을 통해 과거의 구매내역이나 추천 결과를 쉽게 반영한다는 장점이 있다. 그러나 서로 다른 아이템이라 하더라도 특징을 나타내는 내용 집합이 동일한 상태일 때 두 아이템의 구분을 못한다는 단점이 있다[10]. 항목의 내용 정보가 풍부하고 분석하기 용이한 텍스트 문서, 뉴스, 웹 페이지 등의 항목 추천에 적합하다[11]. 사용자간의 선호 정보를 활용하는 것이 아니라 각 사용자별 특정 콘텐츠 또는 상품의 속성을 학습하는 것을 기반으로 한다.

2.2.3 인구통계학적기반 필터링(Demographic-based filtering)

나이, 성별, 교육 수준, 직업 등의 사용자 정보를 활용하여 추천을 수행하는 시스템이다[12]. 인구 통계학적 속성은 다양한 종류의 항목이나 항목 범주에 대한 사용자의 성향을 분석하기 쉬우며 구매 이력과 같은 데이터가 없는 초기 단계에

있는 사용자의 개인 성향을 분석하고 추천하는 것이 가능하다.

2.2.4 협업 필터링(collaborative filtering)

협업 필터링은(Collaborative Filtering, CF)은 많은 사용자들로부터 얻은 기호 정보(taste information)에 따라 사용자들의 관심사들을 자동적으로 예측하게 해 주는 방법으로 개인화 추천 시스템에 가장 많이 사용된다. 협력 필터링 접근법의 근본적인 가정은 사용자들의 과거의 경향이 미래에서도 그대로 유지 될 것이라는 전제에 있다[13]. 즉, 고객들의 취향은 무작위로 분포하는 것이 아니라 일반적인 경향과 패턴이 있다는 가정에서 시작된다. 비슷한 취향을 가진 고객들에게 서로 아직 구매하지 않은 상품들을 교차 추천하거나 분류된 고객의 취향이나 생활 형태에 따라 관련 상품을 추천하는 형태의 서비스를 제공하기 위해 사용된다. 하지만 충분한 데이터가 확보되어야지만 좋은 결과를 나타내고, 특별한 성향에 대하여는 예측력이 떨어지고 유사한 데이터가 많을수록 많은 계산량을 요구하는 단점이 있다[14][15].

협업필터링은 3단계를 걸친다[16].

제1단계“ 평가치 매트릭스의 준비

p명의 고객들이 q개의 항목을 경험하고 부여한 평가치들을 정리하여 [표2-1]과 같이 p x q의 고객 x 항목 매트릭스를 만든다.

[표 2-1 협업필터링에서 고객 x 항목 매트릭스]

고객 \ 아이템	아이템 1	아이템 2	아이템 3	아이템 4
고객 A	$R_{A,1}$	\emptyset	$R_{A,3}$	$R_{A,4}$
고객 B	$R_{B,1}$	$R_{B,2}$	\emptyset	$R_{B,4}$
고객 C	$R_{C,1}$	\emptyset	$R_{C,1}$	\emptyset

여기서 $R_{A,1}$ 는 고객 A가 아이템 1에 부여한 평가치를 뜻한다. \emptyset 는 고객이 아이템을 평가하지 않았음을 뜻한다.

제 2단계 : 최근접 이웃의 구성

고객들의 평가치를 이용하여 고객들 간의 유사도를 계산한 후 최근접 이웃을 구성한다. 유사도를 계산하기 위한 측정지수로는 일반적으로 Pearson Correlation Coefficient와 Cosine이 사용된다.

Pearson Correlation Coefficient $\omega(A,B)$ 는 두 고객 A, B에 의해 공통적으로 평가된 아이템들의 평가치를 이용하여 다음과 같이 계산한다.

$$\omega(A,B) = \frac{\sum_{i=1}^q (R_{A,i} - \overline{R_A})(R_{B,i} - \overline{R_B})}{\sqrt{\sum_{i=1}^q (R_{A,i} - \overline{R_A})^2 \sum_{i=1}^q (R_{B,i} - \overline{R_B})^2}}$$

여기서 $\overline{R_j}$ 는 고객 A의 최근접 이웃인 고객 j의 이용가능한 평가치들의 평균 값을 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.

여기서 $R_{A,i}$ 와 $R_{B,i}$ 는 고객 A와 B가 공통으로 평가한 아이템 I의 평가치를 뜻한다. 그리고 $\overline{R_A}$, $\overline{R_B}$ 는 고객 A와 B의 이용 가능한 평가치들의 평균 값을 뜻한다.

Cosine $\cos(\vec{A}, \vec{B})$ 는 고객 A와 B가 공통으로 평가한 아이템들의 평가치를 q차원 공간에 벡터화한 후 두 벡터 사이 각의 Cosine 값을 다음과 같이 계산한다[15].

$$\cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\|^2 \|\vec{B}\|^2}$$

Herlocker et al.(1999)[16]은 고객들 간의 유사도를 계산할 때, Pearson Correlation Coefficient를 사용하는 것이 Cosine을 사용하는 것보다 높은 추천 성과를 보인다고 발표하였다. 따라서 본 논문에서는 최근접 이웃을 구성하기 위해 Pearson Correlation Coefficient를 사용하였다.

목표고객에 대한 다른 고객들의 유사도를 계산한 후 목표고객의 최근접 이웃을 다음의 두 가지 방법으로 구성할 수 있다. 첫 번째는 목표고객과의 유사도가 미리 설정한 Threshold값 이상인 고객들로 최근접 이웃을 구성하는 것이고, 두 번째는 K-NN(K-Nearest Neighbors) 기법으로서 목표고객과의 유사도가 높은 상위 k명의 고객들로 최근접 이웃을 구성하는 것이다.

제 3단계 : 추천목록의 생성

최근접 이웃의 평가치를 이용하여 목표고객이 평가하지 않은 아이템의 평가치를 예측한 후 추천목록을 생성한다. 평가치 $R_{A,i}$ 는 고객 A의 아이템 I에 대한 평가치로 고객 A의 최근접 이웃의 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 A의 최근접 이웃인 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.

협업필터링은 위와 같은 단계를 거치면서 고객에 대한 추천목록을 생성한다. 협업 필터링은 대부분의 추천 시스템의 주요한 데이터 마이닝 기법으로 사용된다. 하지만 이러한 협업필터링은 희박성(Sparsity), 확장성(Scalability) 그리고 투명성(Transparency) 등의 한계점을 가지고 있다.

희박성(Sparsity)은 고객이 아이템에 부여한 평가치의 개수가 부족하여 추천의 성과가 떨어지는 문제점이다. 취급하는 아이템의 종류가 많은 Amazon이나 e-Bay 같은 대형 인터넷 쇼핑몰에서는 고객들이 실제로 경험하고 평가하는 아이템들의 개수가 전체의 1%도 안될 뿐만 아니라 평가를 하지 않는 고객들도 많다. 즉, 협업 필터링 과정 중 제 1단계에서 준비되어야 하는 고객 × 아이템 매트릭스

의 대부분이 비어 있게 된다. 그러므로 협업필터링이 제대로 작동할 수 없게 되는 것이다. 희박성의 문제를 완화하기 많은 연구들이 진행되고 있다. Breese(1998)는[17] 평가치가 비어 있는 칸에 Default 평가치를 부여하는 방법인 Default Voting을 사용하였고, Sarwar et al.(2001)[18]은 고객들 간이 아닌 아이템들 간의 유사도를 계산하여 추천에 활용하는 방법을 제안하였다. 다른 방법으로는 차원 감소법이 있다. 이는 직접적으로 고객 \times 아이템 매트릭스의 차원을 감소시키는 방법으로서 고객이나 아이템을 군집화하여 군집화 된 그룹을 기본단위로 하여 협업 필터링 기법을 적용하는 방법이다. 그리고 협업 필터링과 내용기반 필터링을 결합한 하이브리드 기법으로 희박성을 완화시키려는 연구가 많이 진행되고 있다.

확장성(Scalability)은 고객의 수와 거래데이터의 개수가 늘어남에 따라 목표고객의 최근접 이웃을 찾기 위한 연산이 기하급수적으로 늘어난다는 문제점이다. 예를 들어, 협업 필터링이 만 명의 목표고객들에게 추천을 하고자 한다면, 만 명의 목표고객들의 최근접 이웃을 찾기 위해 수십 만 명의 고객데이터와 이들이 부여한 수 백만 개의 평가치를 검색해야 한다. 추천시스템이 상대하는 고객의 수나 취급하는 아이템의 개수가 늘어나면 최근접 이웃을 구성하기 위한 계산의 양이 기하급수적으로 늘어나서 추천목록을 생성하기까지 오랜 시간이 걸려서 시스템의 효율성이 떨어지게 된다. 확장성의 문제를 완화하기 위해 개별 고객들 간의 유사도를 계산하는 대신, 아이템들 간의 유사도를 계산하거나 고객을 군집화한 후 유사도를 계산하는 방법들이 제안되었다.

투명성(Transparency)은 추천 결과와 고객의 선호도와 의 관계가 불명확하다는 문제점이다. 협업필터링은 수식계산을 포함한 블랙박스를 통하여 고객에게 추천 목록을 제시하기 때문에, 고객이 아이템들이 어떻게 추천 되었는지를 이해하는 것이 어려울 수 있다.

협업 필터링은 여러 분야에서 널리 연구 및 적용되고 있다. 기사 추천시스템인 GroupLens[19] 디오 추천 시스템인 Video Recommender[20] 음악 추천 시스템인 Ringo[21] 및 World Wide Web에서 사용자와 관련된 정보를 찾아주는 PHOAKS[22] 등이 개발되어 있으며, Amazon, CDNow, Drugstore, MovieFinder

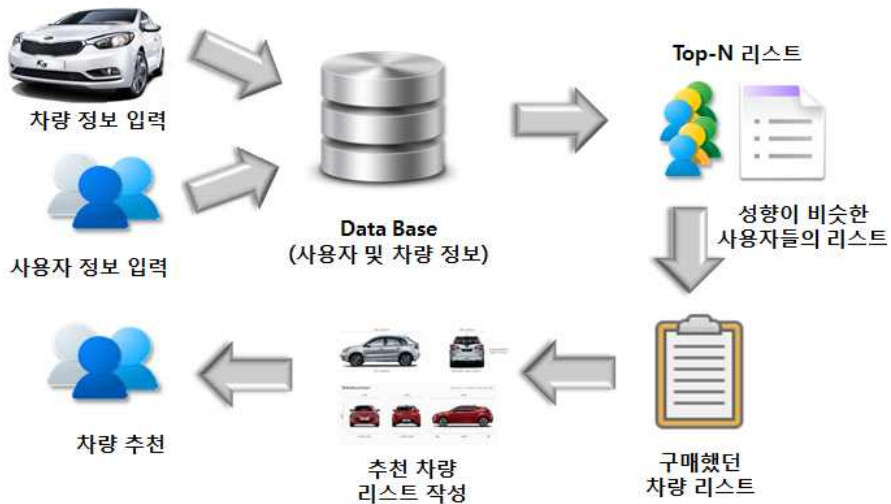
등에서 협업 필터링 기법을 사용하여 상품을 추천하고 있다[23].

제 3 장 협업필터링을 이용한 차량 추천 시스템 설계

본 장에서는 협업 필터링과 개인 성향을 이용한 차량 추천 시스템에 대하여 설계하고, 설계를 위해 전체 시스템의 구조와 차량 추천 시스템의 구성도, 희박성 문제 해결에 대하여 설명한다.

3.1 협업 필터링과 개인 성향을 이용한 차량 추천 시스템 구성도

<그림3-1>과 같은 추천 방법으로 차량 추천 시스템을 구축하고 설계하였다.

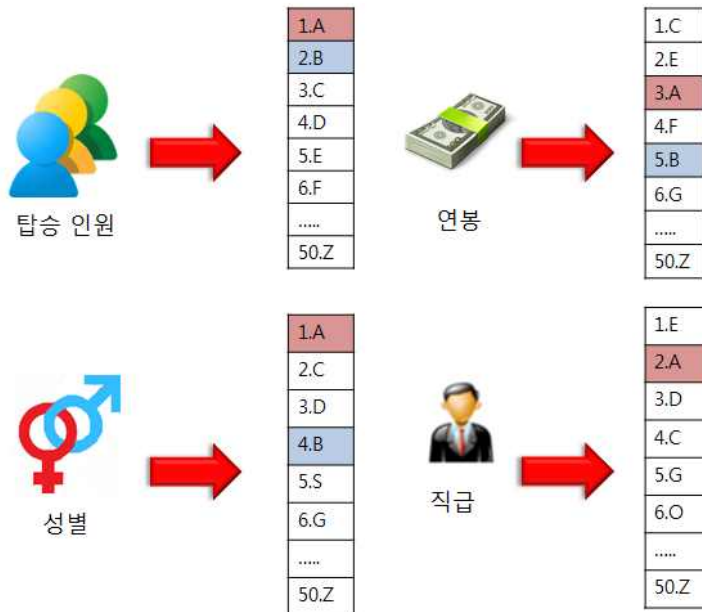


<그림 3-1 시스템 구조도>

우선 데이터베이스에 사용자의 및 차량의 정보를 입력한다. 이때, 입력하는

차량의 정보에는 차량의 이름, 가격, 탑승 가능 인원수, 그리고 차량의 종류이다. 이와 같이 차량의 정보를 입력 할 때 분리한 이유는, 차량을 추천해 줄때에 사용자 의 개인성향에 맞추어 차량을 추천해 주기 위해서이다.

차량을 추천해 줄 때 <그림 3-2>와 같이 형태를 분류 한다.

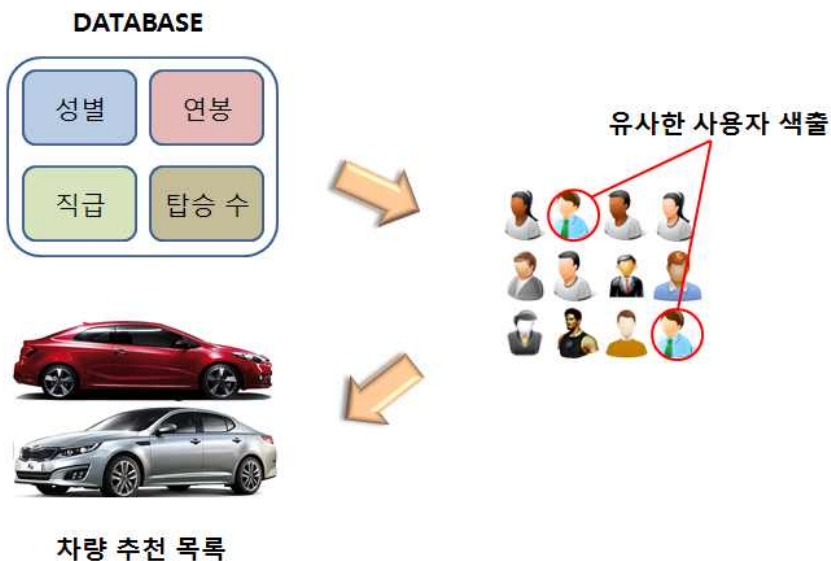


<그림 3-2 차량 추천 방법>

<그림 3 - 2>와 같이 차량을 추천해 줄 때 처음 이용자가 회원가입 할 때에 가입한 탑승 인원, 연봉, 성별, 직급을 가지고 각 카테고리에 따라 차량 구입을 많이 한 차를 1등부터 마지막까지 선정한다. 이렇게 나온 리스트들을 4가지 카테고리에 따라 순위를 비교하였을 때 상위권에 많이 링크된 차량부터 사용자에게 추천할 수 있도록 한다. 위에 그림을 보면 A라는 차량이 4가지 속성에 대하여 모두 상위권에 링크되어 있고 그 다음은 B 자동차가 3가지 속성에 대하여 상위권에 링크되어 있음을 알 수 있다. 그렇기 때문에 추천되는 차량은 A 자동차와 B자동차가 되는 것이다.

3.2 희박성 문제 해결 및 추천 과정

데이터가 없는 신규 회원의 경우 유사도를 이용한 그룹화를 할 수 없는 희박성의 문제가 발생한다. 따라서 회원가입 시 입력 받는 성별, 연봉, 탑승 수, 직급 카테고리의 데이터를 기반으로 사용자와 가장 유사한 사용자들을 찾아 그 유사한 사용자들이 구입한 차량을 추천한다. 다음 <그림 3-2>는 유사 사용자 선정에 대한 개념도를 보여준다. 성별, 연봉, 탑승 수, 직급 카테고리의 데이터를 기반으로 사용자와 가장 유사한 사용자를 찾아 분석하여 그 사용자가 구입한 차량을 이용하여 사용자에게 추천해 준다.



<그림 3-3 유사 사용자 선정 개념도>



DATABASE내의 자료들

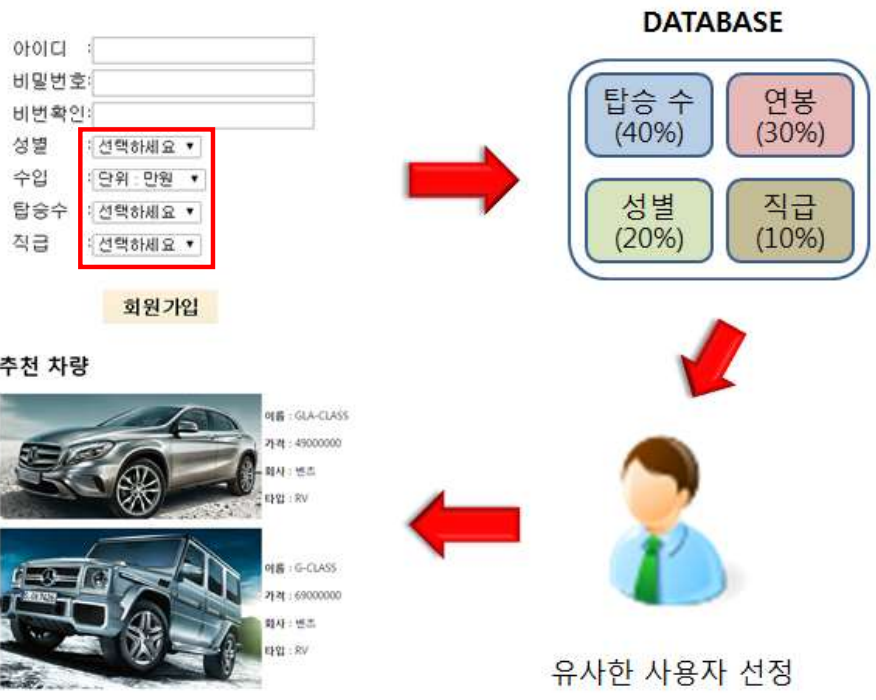
<그림 3-4 유사그룹 형성 과정>

<그림 3-3>는 유사그룹 형성과정으로 협업 필터링의 희박성을 해결하기 위해 회원 가입 시 입력한 정보를 기반으로 유사한 사용자를 선별한다. 이는 회원 가입 시 받은 성별, 연봉, 탑승 인원, 직급등의 카테고리에 각각의 가중치를 다르게 적용 하여, 성향이 유사한 유저를 선별하는 것이다.

[표 3-1 각 카테고리별 가중치]

	가중치
탑승인원	40%
연봉	30%
성별	20%
직급	10%

[표3-1]을 보면 각 카테고리별 적용한 가중치를 확인 할 수 있다. 위 표를 보면 알 수 있듯이 탑승 인원에 가장 많은 가중치를 준 것을 보면 알 수 있는데 이는 만약 탑승 인원의 수가 5인 이상일 경우 탑승 가능 인원이 5인 이상인 RV 종류의 차량을 우선적으로 추천해 주기 위해서 이다. 이를 그림으로 설명한 것이 <그림 3-5>이다.



<그림 3-5 희박성 문제 해결 과정>

제 4장 협업 필터링과 개인 성향을 이용한 차량 추천 시스템 구현

본 장에서는 차량 추천 시스템의 구현환경에 대해 기술하고, 실제 구동의 예와 간단한 설명을 한다.

4.1 구현 환경

[표 4-1]은 전체 시스템의 구현 환경을 나타낸 것이다. 운영체제는 Windows7을 이용하였고, 웹 서버는 개인 PC를 이용하였다. 개발언어는 PHP와 HTML을 이용 했으며, Database는 MySQL을 사용하였다.

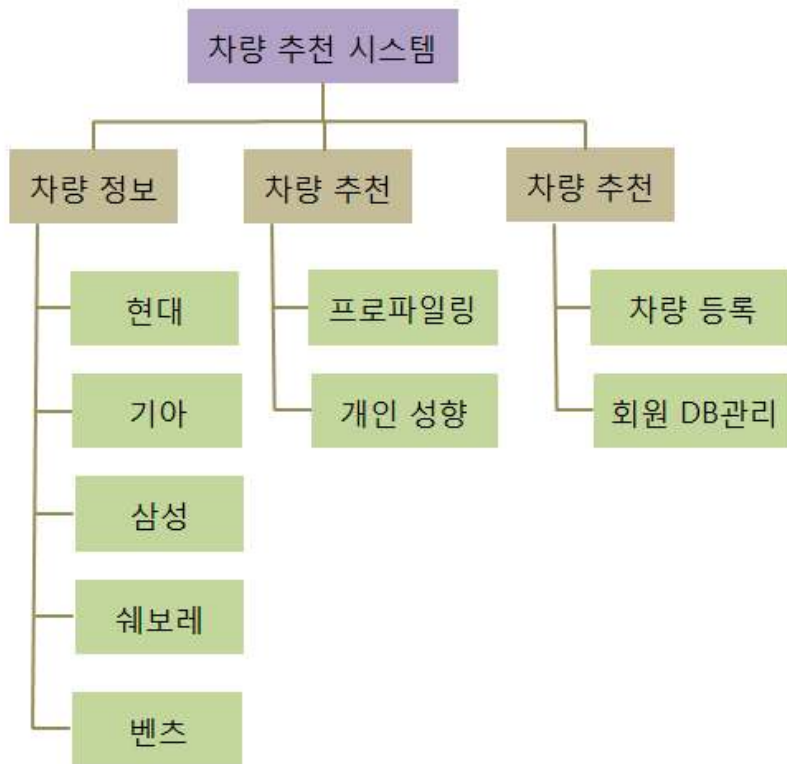
[표 4-1 시스템 구현 환경]

시스템 구성요소	시스템 세부 내용
운영체제	Windows7
개발언어	PHP, HTML
Web Server	개인 PC
Web Language	PHP
DBMS	MySQL

4.2 협업 필터링과 개인 성향을 이용한 차량 추천 시스템 구현 예

차량 추천 시스템의 전체 구성도는 아래 <그림4-1>과 같다. 메인메뉴에서 차량의 정보를 회사별로 선택하여 확인할 수 있고, 추천 서비스는 회원가입 후 로

그인 하여 이용할 수 있다.



<그림 4-1 차량 추천 시스템 구성도>

<그림 4-2>, <그림 4-3>, <그림 4-4>는 각 회사별 차량정보 페이지이다. 메인 메뉴에서 특정 회사를 선택하게 되면 이름을 기준으로 오름차순으로 정렬되어 보여준다.



이름 : K3
가격 : 19,300,000원



이름 : K3 coupé
가격 : 18,300,000원



이름 : K5
가격 : 20,400,000원



이름 : 모닝
가격 : 87,100,000원



이름 : 모하비
가격 : 38,360,000원



이름 : 스포티지
가격 : 27,680,000원

<그림 4-2 기아 차량 정보화면>



이름 : I30
가격 : 18,200,000원



이름 : I40
가격 : 27,500,000원



이름 : 그랜저
가격 : 30,240,000원



이름 : 벨로스터
가격 : 38,300,000원



이름 : 벨로스터
가격 : 18,100,000원



이름 : 싼타페
가격 : 20,800,000원

<그림 4-3 현대 차량 정보화면>



이름 : ON3
가격 : 22,500,000원



이름 : ON5
가격 : 26,700,000원



이름 : SM3
가격 : 15,760,000원



이름 : SM32.E
가격 : 46,200,000원



이름 : SM5
가격 : 22,300,000원



이름 : SM5 D
가격 : 25,800,000원

<그림 4-4 삼성 차량 정보화면>

<그림 4-5>은 회원가입 페이지이다. 회원 가입 시 성별, 수입, 탑승 수, 직급 등을 입력하여 회원의 개인 성향을 분석한다.

아이디 :

비밀번호:

비번확인:

성별 : ▼

수입 : ▼

탑승수 : ▼

직급 : ▼

회원가입

<그림 4-5 회원 가입 시 입력할 정보>

<그림 4-6> 차량이 추천 되는 모습을 볼 수 있다.

순천향대학교

HYUNDAI

KIA MOTORS

**SAMSUNG
RENAULT SAMSUNG
MOTORS**

CHEVROLET

Mercedes-Benz

Park님 환영합니다

로그아웃 회원가입

추천받기

추천 차량

이름 : GLA-CLASS
 가격 : 49000000
 회사 : 벤츠
 타입 : RV

이름 : G-CLASS
 가격 : 69000000
 회사 : 벤츠
 타입 : RV

순천향대학교

HYUNDAI

KIA MOTORS

**SAMSUNG
RENAULT SAMSUNG
MOTORS**

CHEVROLET

Mercedes-Benz

moon님 환영합니다

로그아웃 회원가입

추천받기

추천 차량

이름 : 모닝
 가격 : 8710000
 회사 : 기아
 타입 : sedan

이름 : 엑센트
 가격 : 11350000
 회사 : 현대
 타입 : sedan

<그림 4-6 이용자에 따라 다른 차량이 추천>

제 5장 결론 및 향후 연구과제

개인의 수입이 증가됨에 따라 차량에 대한 관심이 많아지고 또, 수많은 종류와, 기능을 가진 차량들이 증가함에 따라 사용자에게 맞는 서비스를 제공하는 것이 중요하다 생각한다. 이에 추천 서비스를 제공하여 사용자에게 맞는 차량을 추천해 주는 시스템을 구축 하였다.

본 논문에서 설계하고 구현한 차량 추천 시스템은 협업필터링을 기반 차량 추천 시스템의 연구 내용을 요약하면 다음과 같다.

첫째, 추천 시스템에서 가장 많이 쓰이는 협업필터링을 이용해 고객에게 추천을 했다.

둘째, 데이터의 희박성을 개선시키고자 고객에게 입력 받은 데이터를 토대로 가중치를 두는 프로파일링 기법을 적용하여 추천했다.

본 시스템을 운영하기 위해서는 무수히 많은 차량의 정보를 가지고 있어야 하며 이를 데이터베이스에 입력하여야 한다. 하지만 DB의 정보가 극히 제한적이며, 대량의 정보를 수집한다 하여도 데이터 입력 절차에 한계가 있기에 기본적으로 필요한 데이터만으로 구축해 놓았다. 또한 분석을 바탕으로 하는 시스템은 대량의 데이터를 기반으로 실행되어야 하는데 이러한 데이터를 구하여 높은 정확성과 신뢰성을 위하여 Big-Data에 대한 이해와 적용으로 더욱 발전된 시스템을 구축해야 할 것이다.

참 고 문 헌

- [1] M. Bamshad, R. cooley, J. Srivastava, Web Mining : Information and Pattern Discovery on the World Wid Web, In Proceedings of the 9th IEEE International Conference on Tools with Arificial Intelliigence(ICTAI'97), November 1997

- [2] M.S, Chen, J. S. Park, P.S.Yu., Data Mining for path traversal patterns in a Web environment, In proc. 1th International Conference on Distrubuted Computing Systems, pp385-392, 1996

- [3] Wikipedia, Wikimedia Foundation Inc. <http://ko.wikipedia.org/>

- [4] P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl, "GroupLens: and open architecture for collaborative filtering of Netnews." In Proceedings of Computer Supported Cooperative Work Conference, 2001, pp. 175-186

- [5] 갈리트 시뮤엘리, 비즈니스 인텔리전스를 위한 데이터마이닝, 사이텍미디어, 2009

- [6] Hill, W., L. Stead, M. Rosenstein and G. Furnas, "Recommending and evaluating choices in a virtual community of use," In proceeding of CHI 95, 1995.

- [7] 김재경, 서지혜, 안도현, 조윤희, "협업 필터링 기법을 활용한 개인화된 상품 추천방법론 개발에 관한 연구", 한국지능정보시스템학회, 제 8권, 제2호, 2002, pp139-157

- [8] Moody, James and DoubbMody, James and Douglas R. White,"Structural Cohesion and Embedednes: A Hierarchical Concept of Social Groups", American Sociological Review, Vol.68, 203, p.103-127.

- [9] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating 'word of mouth'," In Proceedings of ACM CHI'95 1995
- [10] Loren Terveen, Will Hill, Brian Amento, David McDonald, Josh Creter, "PHOAKS: a system for sharing recommendations," Communications of the ACM Vol.40, No.3, 1997, pp. 59-62.
- [11] 강의영, "Development and Implementation of a Frame work for Personalized Recommender System based on Htbrid Filtering",
- [12] 안신현, "Issues of contents Recommendation System", KAIST문화기술대, 2009
- [13] Yang, W., Z. Weng and M. You, "An Improved Collaborative Filtering Method for Recommendations Generation," IEEE Int'lConf. on Systems, Man and Cybernetics(2004), 4135~4139
- [14] 이세일 외, "A Recommendation System using content-based Collavoration Large Data", 정보과학회논문지: 컴퓨팅의 실제 및 레터, 제 18권, 제 1호, p55~59,2012.01
- [15] Sarwar et al. "Hybrid recommendation systems that take benefits from both collaborative filtering and content-based learning have been also found in the literature", 2000
- [15] 이재식, 박석두 "장르별 협업필터링을 이용한 영화추천시스템 성능향상" 한국 지능정보시스템학회논문지 제 13권 제4호 2007년 12월(pp.65~78)
- [16] Sarwar et al. "Hybrid recommendation systems that take benefits from both collaborative filtering and content-based learning have been also found in the literature", 2000

- [17] Rosa M. Rodriguez, Macarena Espinilla, Pedro J. Sanchez, "Using linguistic incomplete preference relations to cold start recommendations", *Internet Research* Vol20, issue 3 pp.296-315
- [18] Horvitz, Bressan et al , "The Lumière Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users" - n Proceedings of Computer Supported Cooperative Work Conference, 1998, pp. 130-142
- [19] P. Resnick, N. Jacobson, M. Sushak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of Netnews." In Proceedings of Computer Supported Cooperative Work Conference, 2001, pp. 175-186
- [20] Hill, W., L. Stead, M. Rosenstein and G. Furnas, "Recommending and evaluating choices in a virtual community of use," In proceeding of CHI 95, 1995.
- [21] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating 'word of mouth'," In Proceedings of ACM CHI'95 1995
- [22] Loren Terveen, Will Hill, Brian Amento, David McDonald, Josh Creter, "PHOAKS: a system for sharing recommendations," *Communications of the ACM* Vol.40, No.3, 1997, pp. 59-62.
- [23] Woon-hae Jeong, Se-jun Kim, Doo-soon Park and Jin Kwak, "Performance Improvement of a Movie Recommendation System based on Personal Propensity and Secure Collaborative Filtering," *Journal of Information Processing Systems* Volume 9, Number 1, 2013

[부 록] 차량 추천 시스템 프로그램

insert.php

```
<!DOCTYPE HTML>
<html>

<body>

<table>
<tr>
        <td><a href="main.php?id=<?echo "$id";?>"></a></td>
        <td></td>
        <td><a href="login.php" target="recommend"></a></td>
        <td></td>
</tr>
</table>
<table border="0" cellspacing="0" width="1254" height="107" align="center"
bordercolordark="white" bordercolorlight="black">
<tr>
<td width="205" height="105">
        <p><a href="hyundae.html" target="recommend"></a></p>
</td>
<td width="205" height="105">
        <p><p><a href="kia.html" target="recommend"></a></p>
</td>
<td width="205" height="105">
        <p><p><a href="samsung.html" target="recommend"></a></p>
</td>
<td width="205" height="105">
        <p><p><a href="chevrolet.html" target="recommend"></a></p>
</td>
<td width="205" height="105">
        <p><p><a href="Benz.html" target="recommend"> 과장급이하 <option value="high"> 부장급이상 </select></div>

<div style= "position:absolute; top: 370px; left: 150px; width:115px; height:62px;"></div>
</div>

<div style="position:absolute;top:700px;width:1200px;height:62px;">
<table align="left" style="margin-left:10px;"><tr>
    <td style="padding-right:15px"> </td>
    <td valign="top"> <div class=copyright>
        충청남도 아산시 신창면 읍내리 순천향대학교 멀티미디어관 515호<br />
        Copyright &copy; <a
href="http://comp-cse.sch.ac.kr/cosmetics" target="_blank" title="새창"><strong><font
color="blue">Programming Language Lab</font></strong></a>. <span>All Rights
Reserved.</span>
    </div> </td>
</tr> </table>

</div>

</body>
</html>

```

insert_member.php

```

<?php
//header("Content-Type: text/html; charset=UTF-8");
$connect = mysql_connect("127.0.0.1", "root", "apmsetup");
mysql_selectdb("car");
mysql_query("set names utf8");

// id,pw,성별,수입,탑승인원수,직급을 데이터베이스에 저장하는 부분
$id = $_POST[id];
$pw = $_POST[pw1];
$pw2 = $_POST[pw2];
$sex = $_POST[sex];
$income = $_POST[income];

if($income=="25") $inc="2000~5000";
else if($income=="510") $inc="5000~10000";
else $inc="10000~";

$number = $_POST[number];

```

```

$grade = $_POST[grade];

$qry = "select * from member_data;";

$result = mysql_query($qry);

if($id==" || $pw==" || $pw2==" || $sex==" || $income==" || $number==" ||
$grade=="")
{
    ?>
    <script type = "text/javascript">alert("빈칸이 있습니다. 다시 입력해주세요.")</script>
    <?
        echo ("<script>location.href='login.php'</script>");
        exit;
}

while($obj = mysql_fetch_object($result))
{
    if($id==$obj->mb_id)
    {
        ?>
        <script type = "text/javascript">alert("중복된 아이디 입니다")</script>
        <?
            echo ("<script>location.href='login.php'</script>");
            exit;
        }
    }

if($pw == $pw2){}
else{
    ?>
        <script type = "text/javascript">alert("비밀번호가 같지 않습니다.")</script>
        <?
            echo ("<script>location.href='login.php'</script>");
            exit;
        }
}

$qry = "insert into member_data (id, pw, sex, income, number, grade) values('$id',
'$pw', '$sex', '$inc', '$number', '$grade')";
?>

```

```

<script type = "text/javascript">alert("회원가입완료")</script>
<?
echo ("<script>location.href='login.php'</script>");

$result = mysql_query($qry);

@mkdir($dir);

echo $value;
?>

```

login.php

```

<html>
  <head>
    <meta charset = utf-8>
  </head>
  <body>
<table>
<tr>
  <td><a href="main.php?id=<?echo "$id";?>"></a></td>
  <td></td>
  <td><a href="login.php" target="recommend"></a></td>
  <td></td>
</tr>
</table>

<table border="0" cellspacing="0" width="1254" height="107" align="center"
bordercolordark="white" bordercolorlight="black">
  <tr>
    <td width="205" height="105">
      <p><a href="hyundae.html" target="recommend"></a></p>
    </td>
    <td width="205" height="105">

```



```

<div style="position:absolute;top:187px;left:54px;"> · 비밀번호:<input type="password"
name="pw" style="width:150px;height:17px;"></div>

<!-- <div style="position: absolute; top:152px; left: 240px; width: 115px; height:
62px;"> <input type ="submit" value="?"> --> <div style= "position: absolute; top:
152px; left: 280px; width:115px;height:124px;"></div>

<div style="position: absolute; top:215px; left:54px;"><img src = "img/search_down.gif"
width=340></div>

<div style="position:absolute;top:250px;left:32px;width:225px;height:62px;"> · 아직 회원이
아니십니까?</div>

<div style="position:absolute; top:250px; left:280px; width:115px; height:62px;"> <a href
="insert.php"> </div>
</div>

<div style="position:absolute;top:700px;width:1200px;height:62px;">

<table align="left" style="margin-left:10px;"><tr>
    <td style="padding-right:15px"> </td>
    <td valign="top"> <div class=copyright>
        충청남도 아산시 신창면 읍내리 순천향대학교 멀티미디어관 515호<br />
        Copyright &copy; <a href="http://comp-cse.sch.ac.kr/cosmetics" target="_blank" title="새
        창"><strong><font color="blue">Programming Language Lab</font></strong></a>. <span>All
        Rights Reserved.</span>
    </div> </td>
</tr> </table>

</div>

</body>

</html>

```

LoginMgr.php

```
<?php
$connect = mysql_connect("127.0.0.1", "root", "apmsetup");
mysql_selectdb("car");
mysql_query("set names utf8");

$id = $_POST[id];
$pw = $_POST[pw];

echo "$id";
echo "$pw";

$qry = "select * from member_data where id='$id'";
$result = mysql_query($qry);
$cnt=mysql_num_rows($result);
if($cnt==0)
{
?>
<script type = "text/javascript">alert("No id in Database")</script>
<?
    echo ("<script>location.href='login.php'</script>"); //메인페이지로 갈부분
    exit;
}
else
{
    $obj = mysql_fetch_object($result);
    if($pw==$obj->pw)
    {
        echo ("<script>location.href='main.php?id=$id'</script>");
        //메인페이지로 갈부분
    }
    else
    ?>
    <script type = "text/javascript">alert("password error")</script>
    <?
    echo ("<script>location.href='login.php'</script>"); //메인페이지로 갈부분
}
?>
```


main.php

```
<html>

<head>
<meta http-equiv="content-type" content="text/html; charset=euc-kr">
<title>자동차 추천 프로그램</title>
<base target="top">

</head>
<p align="center" ></p>
<table border="1" cellspacing="0" width="1254" height="107" align="center"
bordercolordark="white" bordercolorlight="black">
  <tr>
    <td width="205" height="105">
      <p><a href="#" target="_self"></a></p>
    </td>

    <td width="205" height="105">
      <p><p><a href="#" target="_self"></a></p>
    </td>

    <td width="205" height="105">
      <p><p><a href="#" target="_self"> </a></p>
    </td>

    <td width="205" height="105">
      <p><p>
      <a href="#" target="_self"> </a></p>
    </td>

    <td width="205" height="105">
      <p><p><a href="#" target="_self"></a></p>
    </td> <td width="203" height="105">
```

```

                <p><p><a href="#" target="_self"></a></p>
            </td>
        </tr>
    </table>

<p>&nbsp;</p>
</body>

</html>

```

index.php

```

<html>
<head><meta http-equiv="Content-Type" content="text/html; charset=utf-8"></head>
<body><font face="맑은 고딕"> <h1> 추천 차량</h1><table><?
    $connect = @mysql_connect("127.0.0.1","root","apmsetup") or die("DB connect
error");
    $connect = @mysql_select_db('car',$connect) or die("DB select error");

    mysql_query("set session character_set_connection=utf8;");
    mysql_query("set session character_set_results=utf8;");
    mysql_query("set session character_set_client=utf8;");

    $mb_id = $_GET[id];

    if($mb_id=="")
    {
        ?>

        <?
        echo ("<script>location.href='main.php'</script>");
        exit;
    }
    else{

        // 로그인한 회원의 정보를 데이터베이스에서 가져옴
        $sql = "select * from member_data where id='$mb_id'";
        $select_result = mysql_query($sql);
        for($i=0;$compare = mysql_fetch_array($select_result);$i++)
        {
            $mb_sex = $compare[sex];

```

```

        $mb_income = $compare[income];
        $mb_number = $compare[number];
        $mb_grade = $compare[grade];
    }

    $count = 0;

    $sql = " delete from recommend ";
    $result = mysql_query($sql);

    $sql = "select * from member";
    $result = mysql_query($sql);

    // Top-N-List를 구성하는 알고리즘
    // 기존의 회원정보들과 비교해 최대3점 최소 1점
    // 40% : 탑승인원수
    // 30% : 연봉
    // 20% : 성별
    // 10% : 직급

    for($i=0;$row=mysql_fetch_array($result); $i++)
    {
        if($row[sex]==$mb_sex)
            $total=0.2*3;
        else
            $total=0.2*1;

        if($row[income]==$mb_income)
            $total=$total+(0.3*3);
        else
            $total=$total+(0.3*1);

        if($row[number]==$mb_number)
            $total=$total+(0.4*3);
        else
            $total=$total+(0.4*1);

        if($row[grade]==$mb_grade)
            $total=$total+(0.1*3);
        else
            $total=$total+(0.1*1);
    }

```

```

// 데이터베이스에 Top-N-List를 저장하는 부분
        $sql = " insert into recommend
set login_id = '$mb_id',
        id = '$row[id]',
        total = '$total' ";

mysql_query($sql);

}

// 성향이 가장 비슷한 2명의 정보를 바탕으로 차량을 추천해주는 알고리즘
$sql = "select * from recommend order by total desc limit 0,5";
$result = mysql_query($sql);
for ($i=0; $row=mysql_fetch_array($result); $i++)
{
    $sql_select = "select distinct car_id,sex,income,number,grade,id from
member where id=$row[id]";
    $test = mysql_query($sql_select);
    for($j=0;$row2=mysql_fetch_array($test);$j++)
    {
        $sql_select2 = "select * from car_info where id=$row2[car_id]
order by id desc limit 0,3";
        $test2 = mysql_query($sql_select2);
        for($n=0;$row3=mysql_fetch_array($test2);$n++)
        {
            if($count<2)
            {
                ?><tr><td align=center size=800>.png"</td>

<td>
<b>이름</b> : <?echo "$row3[name] <br><br>"; ?>

<b>가격</b> : <?echo "$row3[price] <br><br>"; ?>

<b>회사</b> : <?echo "$row3[corp] <br><br>"; ?>

<b>타입</b> : <?echo "$row3[ type] <br>"; ?>

</td>
<?
$count = $count + 1;
}

```

```
        }
        ?>
</tr><tr></tr><tr></tr><tr></tr><tr></tr><tr></tr>
        <?
        }
    }
?>
</table>
</font>
</body>
</html>
```