

# 시맨틱 웹에서 다중 혼합필터링을 이용한 개인화된 의상 코디 시스템

은채수<sup>○</sup> 송창우 이승근 이정현  
인하대학교 컴퓨터정보공학과

{chaesoo79<sup>○</sup>, whrma, sglee}@hci.inha.ac.kr, jhlee@inha.ac.kr

## Personalized Apparel Coordi System

### using Multiple Hybrid-Filtering on Semantic Web

Chaesoo Eun<sup>○</sup> Changwoo Song Seunggeun Lee Junghyun Lee  
Computer Science & Engineering, Inha University

#### 요 약

인터넷과 웹이 일상생활의 일부가 되면서 온라인상에는 방대한 양의 정보가 쌓이게 되었다. 이러한 흐름 속에서 정보의 양은 급속도로 늘어나는 현상을 보이며, '개인화'를 통해 수많은 데이터들 사이에서 원하는 정보를 자동으로 찾아내는 기술의 중요성이 부각되고 있다. 이를 '추천시스템'이라 부르며, 내용기반 필터링과 협력적 필터링 등의 연구가 활발히 이루어지고 있다.

그러나 사용자에게 가장 중요한 영향을 미치는 도래의 선호도, 지역, 시대 등의 복합적인 환경을 반영하는데 아직까지 어려움을 지니고 있다. 따라서 본 논문에서는 기존의 필터링들을 조합하고 좀더 편리하게 정보를 공유하고 학습할 수 있는 시맨틱 웹에서 연관 이웃 마이닝 기법을 통해 개인화된 추천 시스템을 설계한다. 생활에서 흔히 접할 수 있는 의상을 다양한 사용자에게 특화되어 코디해주는 시스템을 웹에서 제공한 결과 불필요한 검색시간이 줄어들고 사용자의 피드백을 통해 점차 만족도가 향상됨을 알 수 있었다.

## 1. 서 론

오늘날 우리는 디지털기와 정보의 바다에 빠져있다. 인터넷을 통해 전 세계가 연결되어 있는 것이다. 그에 따라 사용자들의 요구가 다양화되고 확실화된 서비스의 수명주기가 단축되는 등의 환경변화에 능동적으로 적응하기 위하여 많은 연구자들이 대량의 정보를 자동화 처리할 수 있는 방법을 찾고 있다[1]. 이러한 변화에 연구자들은 '개인화'를 통해 사용자의 취향과 감각을 지능적으로 분석하여 구매할 콘텐츠를 적절하게 추천하는 특화된 서비스의 제공을 목표로 하고 있다.

이러한 서비스를 제공하기 위해 사용자가 원하는 정보를 신속하고 정확하게 전달할 수 있는 추천시스템[2]이 제안되었다. 이때 정확성을 높이기 위해서 어떤 사용자의 데이터를 활용할 것인가, 어떤 기법을 사용하여 추천을 사용할 것인가에 따라 결과가 크게 달라진다. 대표적으로 내용기반 필터링, 협력적 필터링 등의 추천 방법이 있다. 그러나 전통적인 필터링 기법들은 단독으로 사용되기에는 새로운 사용자 또는 콘텐츠들이 끊임없이 등장하는 상황에서 이를 즉각적으로 반영하기가 어렵다.

최근에는 시맨틱 웹의 등장으로 자동화 기술을 이용해 좀 더 편리하게 정보를 공유하고 학습할 수 있게 되었다[3]. 현재 추천 시스템은 주로 협력적 필터링을 사용하고 있는데, 이것은 특정 사용자와 유사한 사용자 군집을

구성해 이들의 선호도를 추천에 반영하는 방식이다. 그러나 이 방식은 처음 나온 콘텐츠에 대해서는 아무런 추천 기능을 수행할 수 없는 데다, 어느정도 사용자들의 프로파일이 모이기 전에는 자신과 유사한 제품만을 사용한 특정 사람을 찾기가 쉽지 않다는데 문제가 있다. 이를 보완하기 위해 내용기반 필터링등 다양한 방식들을 함께 사용[4]하고 있다.

예를 들어 인터넷 서점인 아마존(Amazon.com)은 비슷한 성향의 사용자들이 이미 구매한 상품을 자동적으로 추천하는 시스템을 갖추고 있다. 아마존이 사용하고 있는 추천 시스템은 백 개가 넘는 변수를 포함한 복잡한 수식으로 계산된다. 변수에는 특정 소비자가 좋아하는 연기자, 작가, 감독, 장르 등이 모두 포함된다[5]. 특히 옷의 경우에는 어느 시대에 어떤 의도로 이를 구현했는가에 따라 삶의 모습과 다양한 문화를 나타내므로 도래의 선호도는 물론 지역, 시대 등의 복합적인 요소를 고려해야 한다.

따라서 본 논문에서는 내용기반 필터링과 협력적 필터링을 적절하게 조합하고 시맨틱 웹에서 해석된 성별, 나이, 직업, 선호도 등의 메타데이터(meta-data)로부터 콘텐츠의 특성과 사용자들의 유사점을 추출하였다. 또한 연관 이웃 마이닝을 사용하여 개인화 시스템에서 중요한 잘못된 정보를 줄임으로써 결과의 질을 향상시켰다.

## 2. 관련연구

이 장에서는 본 논문에서 제안한 시스템을 이해하기 위해서는 시맨틱웹이 무엇이고, 어떻게 마이닝을 이용하여 추천을 할 수 있는지 먼저 알아야한다. 그리고 개인화 추천에 대해서 좀 더 자세히 알아본다.

### 2.1. 시맨틱 웹과 데이터 마이닝

기존의 웹에서의 HTML을 이용한 표현방식은 사용자에게 문서의 내용을 디스플레이 하는 정도로 사람이 아닌 프로그램 또는 소프트웨어 에이전트가 자동으로 문서로부터 의미를 추출하기가 어려웠다. 예를 들면, 흔히 말하는 배가 음식으로서의 배인지 운송수단으로서의 배인지 구별하기가 힘들다. 사용자가 원하는 내용과 무관한 많은 양의 내용으로 인해 사용자들은 웹의 검색을 효율적으로 사용할 수가 없었다. 그래서 표현은 같지만 뜻이 다른 데이터를 본래 의미로 해석하고 인식하며, 표준화된 정보를 제공해 줄 수 있는 환경이 시맨틱 웹이다.

데이터 마이닝 기술은 컴퓨터와 데이터베이스 시스템, 그리고 인터넷의 많은 데이터로부터 빈번하게 발생하는 패턴을 찾아서 유용하게 사용할 수 있도록 도와준다. 데이터 마이닝 기법들에는 속성 지향 귀납 추리 (AOI), 결정트리 (decision tree), 연관규칙 (association rules), 분류규칙 (classification) 등이 있는데, AOI와 ID3를 사용하고 있다. 이 기술들이 사람과 기계에 의해 직관적으로 이해 될 수 있을 뿐 아니라 기계적인 처리 절차를 사회화 할 수 있기 때문에 타당성 검증이 용이하다.

시맨틱 웹을 통해 기계에 의해 정보가 이해되어질 수 있고 데이터 마이닝을 사용함에 따라 지식의 재사용이 가능하게 되며, 패턴들 간에 관계나 패턴을 유기적으로 도출할 수 있게 되었다[6]. 본 논문에서는 시맨틱 웹 상에서의 의미 기반의 유용한 정보를 이끌어 내기 위하여 연관 이웃 마이닝 기법을 응용한다.

### 2.2. 개인화 추천

개인화는 이용자의 프로필에 근거 하여 내용을 선정하고, 사용자에게 새롭거나 다른 이용자가 미처 생각하지 못했던 것을 채택하는 것이다. 즉, 사용자의 특성을 대상으로 하여 이와 관련한 다양한 통계의 분석과 비교를 통해 개별화 할 수 있는 지식 및 규칙을 찾아내는 과정이다. 이러한 개인화를 위해서는 추천 시스템이 필요하다. 현재 추천 시스템은 전자상거래 사이트에서 주로 사용되고 있으며 추천 시스템에서 사용되는 필터링 기법은 규칙 기반 필터링, 학습 에이전트, 내용 기반 필터링, 협력적 필터링이 있다.

규칙 기반 필터링은 고객의 선호를 파악하여 정해진 규칙에 따라 콘텐츠를 특성과 연결시켜 추천을 제공한다[7]. 이 규칙은 전문가에 의해 미리 정해지며, 이러한 방식은 콘텐츠의 단가가 비싸거나 콘텐츠 자체의 특성이 복잡할 경우에 효과적이다.

내용 기반 필터링은 콘텐츠의 속성들에 대한 정보를 시스템이 보유하고 있다가 고객이 그 속성에 해당하는 키워드를 입력할 때 그 키워드에 해당하는 속성을 지닌

콘텐츠들을 추천해주는 것으로 정보검색 분야에서 많이 활용된다[8]. 내용 기반 필터링의 단점으로는 첫째가 콘텐츠의 정보가 풍부하다면 추천의 결과가 상당의 좋지만 그렇지 않으면 추천의 결과의 신뢰성이 떨어진다는 점이고 둘째는 추천되는 콘텐츠가 사용자 자신이 가진 취향에서 벗어나지 못하는 특수화 경향을 보이기 쉽다는 것이다. 마지막으로 사용자가 의사 표현을 얼마나 많이 하느냐에 따라 추천의 질이 달라진다.

협력적 필터링은 사용자로부터 콘텐츠 선호도에 대한 데이터베이스를 구축하고, 새로운 고객이 등장하면 유사한 취향을 가진 이웃들을 데이터베이스로부터 찾아낸다[9]. 그리고 이웃들이 선호하는 콘텐츠를 그 사용자 역시 선호할 것이라는 가정 하에 콘텐츠를 고객에게 추천한다. 따라서 협력적 필터링을 사용하면 위의 내용 기반 필터링의 문제점을 보완할 수 있다. 하지만, 연구와 실무 분야에서 모두 매우 성공적으로 평가되어온 협력적 필터링 기법을 활용한 추천 시스템에 있어서는 기본적으로 자신의 취향과 비슷한 사람이 적을 경우 추천의 질이 떨어지고 추천의 대상이 새로 주어지는 경우 이에 대해 누군가가 평가를 하기 전에는 추천이 이루어지지 않는 문제점을 가지고 있다.

## 3. 선호도를 고려한 의상 코디 시스템

이 장에서는 사용자정보와 선호도를 고려하여 의상을 추천하고 코디하기 위해 상황을 정의하고 내용기반 필터링 및 협력적 필터링 기법과 연관 이웃 마이닝을 적절하게 조합하여 의상 코디 시스템을 구성한다.

### 3.1 의상 콘텐츠 정보 생성

내용기반 필터링은 정보 검색이나 정보 필터링 분야에서 자연스럽게 발전하였다. 따라서 내용기반 필터링은 대부분 콘텐츠를 추천하기 위하여 콘텐츠 정보와 사용자가 요구하는 정보간의 유사도를 측정하고 결과를 순위화하여 보여준다. 이렇게 콘텐츠를 중심으로 분석하여 사용자에게 추천하는 기법이 내용기반 필터링이다.

본 시스템에서는 콘텐츠의 내용으로 의상 콘텐츠 정보 DB를 구축하고 사용자 체형에 맞는 프로필을 사용하여 의상을 추천한다. 내용 정보 데이터베이스를 구축하기 위해 웹문서에서 자동으로 구축하는 방법과 사용자가 정보를 직접 입력하는 두 가지 방법을 사용한다. 자동구축에서는 웹 로봇 에이전트에 의해 웹문서를 추출하여 데이터베이스를 구축한다. 아래 <표 1>은 내용 정보 데이터베이스의 예이다.

<표 1> 내용 정보 데이터베이스 구축 예

의상 정보	내용
의상류	정장
상의/하의	하의
색깔	검정
계절	가을
사이즈	30

3.2 사용자간의 관련성 정보 생성

사용자와 유사한 선호도를 가진 사용자 군집의 선호도에 따라 새로운 사용자가 관심을 가질 것으로 생각되는 콘텐츠를 추천해 주는 기법이 협력적 필터링이다.

본 시스템에서는 사용자의 성별, 연령, 직업, 계절, 지역 등의 정보로 의상 콘텐츠 DB를 구축하고 사용자 체형과 맞는 프로파일을 사용하여 의상을 추천한다. 협력 정보 데이터베이스를 구축하기 위해서는 웹문서에서 자동으로 구축하는 방법을 사용한다. 아래 <표 2>는 협력 정보 데이터베이스의 예이다.

<표 2> 협력 정보 데이터베이스의 예

성별	연령	직업	계절	지역	상위:가디건	하위:청바지	파란색	청색
남	10~20	학생	가을	서울	상의:가디건	하위:청바지	파란색	청색
여	10~20	학생	여름	경기도	상의:나시	하위:치마	노란색	흰색
남	20~30	회사원	봄	전라도	상의:난방	하위:정장바지	검정색	흰색
여	20~30	선생님	가을	경상도	상의:브라우스	하위:정장치마	노란색	갈색
남	30~40	자영업	겨울	강원도	상의:점퍼	하위:폴덴바지	흰색	갈색
여	30~40	주부	여름	제주도	상의:반팔티	하위:반바지	빨간색	흰색

3.3 코디네이션 구성 및 정의

의상 코디네이션 시스템을 위한 코디네이션 구성은 사용자 신체(키, 체형), 색깔, 위/아래 대조로 구성된다. 사용자 성별, 체형, 색깔, 위/아래 대조는 미리 온톨로지로 정의하고 데이터는 전문가로부터 입력 받아 데이터베이스를 구축한다.

<표 3> 코디네이션 구성 및 정의

성별	키	체형	색깔	위/아래 대조
남	큰 키에 마른 체형	흰색을 중심으로 회색등의 모노톤	위에는 짙게, 아래는 밝게	위, 아래
	큰 키에 보통 체형	내추럴톤이나 디프한 톤	모두 밝게	모두 밝게
	큰 키에 살찐 체형	흰색과 검정색	위에는 짙게, 아래는 밝게	위에는 짙게, 아래는 밝게
여	보통 체형에 큰 키	줄무늬 체크무늬	위에는 밝게, 아래는 짙게	위에는 밝게, 아래는 짙게
	보통 체형에 보통 키	모든 색상	위에는 밝게, 아래는 짙게	위에는 밝게, 아래는 짙게
	보통 키에 살찐 체형	흰색과 검정색	위보다 아래, 걸보다 속을 밝게	위보다 아래, 걸보다 속을 밝게

본 혼합 필터링 시스템에서는 상황의 정확한 표현, 다양한 관계 기술 등 명시적 규정을 위하여 시맨틱 웹(Semantic Web)에서 사용되는 온톨로지로 상황을 정의

한다.

3.4 연관 사용자 마이닝

사용자의 선택(트랜잭션)으로부터 연관 사용자를 군집한다. 기본 데이터에서 사용자에게 의해 선호도를 평가한 콘텐츠들을 <표 4>의 사용자 트랜잭션으로 재구성한다.

<표 4> 사용자 트랜잭션

트랜잭션 번호 (평가된 콘텐츠)	추출된 사용자
1 - 청바지	U <sub>1</sub> , U <sub>2</sub> , U <sub>3</sub> , U <sub>4</sub>
2 - 정장상의	U <sub>2</sub> , U <sub>3</sub> , U <sub>1</sub> , U <sub>5</sub> , U <sub>6</sub> , U <sub>7</sub> , U <sub>12</sub> , U <sub>8</sub>
3 - 정장하의	U <sub>9</sub> , U <sub>3</sub> , U <sub>2</sub> , U <sub>10</sub> , U <sub>5</sub> , U <sub>11</sub>
4 - 민소매티	U <sub>13</sub> , U <sub>3</sub> , U <sub>14</sub> , U <sub>15</sub> , U <sub>16</sub> , U <sub>17</sub>
5 - 점퍼	U <sub>18</sub> , U <sub>13</sub> , U <sub>3</sub>
6 - 반팔티	U <sub>13</sub> , U <sub>3</sub> , U <sub>19</sub> , U <sub>20</sub> , U <sub>15</sub>
7 - 후드티	U <sub>21</sub> , U <sub>22</sub>
8 - 면바지	U <sub>23</sub> , U <sub>24</sub> , U <sub>25</sub>

<표 4>의 사용자 트랜잭션으로부터 Apriori 알고리즘으로 연관 규칙을 <표 5>의 방법으로 마이닝한다.

<표 5> 연관 사용자 유추 과정

후보 사용자 집합(C <sub>1</sub> )	U <sub>1(2)</sub> , U <sub>2(3)</sub> , U <sub>3(6)</sub> , U <sub>4(1)</sub> , U <sub>5(2)</sub> , U <sub>6(1)</sub> , U <sub>7(1)</sub> , U <sub>8(1)</sub> , U <sub>9(1)</sub> , U <sub>10(1)</sub> , U <sub>11(1)</sub> , U <sub>12(1)</sub> , U <sub>13(1)</sub> , U <sub>14(1)</sub> , U <sub>15(2)</sub> , U <sub>16(1)</sub> , U <sub>17(1)</sub> , U <sub>18(1)</sub> , U <sub>19(1)</sub> , U <sub>20(1)</sub> , U <sub>21(1)</sub> , U <sub>22(1)</sub> , U <sub>23(1)</sub> , U <sub>24(1)</sub> , U <sub>25(1)</sub>
고빈도 사용자 집합(L <sub>1</sub> )	U <sub>1(2)</sub> , U <sub>2(3)</sub> , U <sub>3(6)</sub> , U <sub>5(2)</sub> , U <sub>13(3)</sub> , U <sub>15(2)</sub>
후보 사용자 집합(C <sub>2</sub> )	(U <sub>1</sub> , U <sub>2</sub> ) <sub>(2)</sub> , (U <sub>1</sub> , U <sub>3</sub> ) <sub>(2)</sub> , (U <sub>1</sub> , U <sub>5</sub> ) <sub>(1)</sub> , (U <sub>1</sub> , U <sub>13</sub> ) <sub>(0)</sub> , (U <sub>1</sub> , U <sub>15</sub> ) <sub>(0)</sub> , (U <sub>2</sub> , U <sub>3</sub> ) <sub>(3)</sub> , (U <sub>2</sub> , U <sub>5</sub> ) <sub>(2)</sub> , (U <sub>2</sub> , U <sub>13</sub> ) <sub>(0)</sub> , (U <sub>2</sub> , U <sub>15</sub> ) <sub>(0)</sub> , (U <sub>3</sub> , U <sub>5</sub> ) <sub>(2)</sub> , (U <sub>3</sub> , U <sub>13</sub> ) <sub>(3)</sub> , (U <sub>3</sub> , U <sub>15</sub> ) <sub>(2)</sub> , (U <sub>5</sub> , U <sub>13</sub> ) <sub>(0)</sub> , (U <sub>5</sub> , U <sub>15</sub> ) <sub>(0)</sub> , (U <sub>13</sub> , U <sub>15</sub> ) <sub>(2)</sub>
고빈도 사용자 집합(L <sub>2</sub> )	(U <sub>1</sub> , U <sub>2</sub> ) <sub>(2)</sub> , (U <sub>1</sub> , U <sub>3</sub> ) <sub>(2)</sub> , (U <sub>2</sub> , U <sub>3</sub> ) <sub>(3)</sub> , (U <sub>2</sub> , U <sub>5</sub> ) <sub>(2)</sub> , (U <sub>3</sub> , U <sub>5</sub> ) <sub>(2)</sub> , (U <sub>3</sub> , U <sub>13</sub> ) <sub>(3)</sub> , (U <sub>3</sub> , U <sub>15</sub> ) <sub>(2)</sub> , (U <sub>13</sub> , U <sub>15</sub> ) <sub>(2)</sub>
후보 사용자 집합(C <sub>3</sub> )	(U <sub>1</sub> , U <sub>2</sub> , U <sub>3</sub> ) <sub>(2)</sub> , (U <sub>1</sub> , U <sub>2</sub> , U <sub>5</sub> ) <sub>(0)</sub> , (U <sub>1</sub> , U <sub>2</sub> , U <sub>13</sub> ) <sub>(0)</sub> , (U <sub>1</sub> , U <sub>2</sub> , U <sub>15</sub> ) <sub>(0)</sub> , (U <sub>1</sub> , U <sub>3</sub> , U <sub>5</sub> ) <sub>(1)</sub> , (U <sub>1</sub> , U <sub>3</sub> , U <sub>13</sub> ) <sub>(0)</sub> , (U <sub>1</sub> , U <sub>3</sub> , U <sub>15</sub> ) <sub>(0)</sub> , (U <sub>2</sub> , U <sub>3</sub> , U <sub>5</sub> ) <sub>(2)</sub> , (U <sub>2</sub> , U <sub>3</sub> , U <sub>13</sub> ) <sub>(0)</sub> , (U <sub>2</sub> , U <sub>3</sub> , U <sub>15</sub> ) <sub>(0)</sub> , (U <sub>2</sub> , U <sub>5</sub> , U <sub>15</sub> ) <sub>(2)</sub> , (U <sub>2</sub> , U <sub>5</sub> , U <sub>13</sub> ) <sub>(0)</sub> , (U <sub>3</sub> , U <sub>5</sub> , U <sub>13</sub> ) <sub>(0)</sub> , (U <sub>3</sub> , U <sub>5</sub> , U <sub>15</sub> ) <sub>(0)</sub> , (U <sub>3</sub> , U <sub>13</sub> , U <sub>15</sub> ) <sub>(2)</sub> , (U <sub>13</sub> , U <sub>15</sub> , U <sub>1</sub> ) <sub>(0)</sub> , (U <sub>13</sub> , U <sub>15</sub> , U <sub>2</sub> ) <sub>(1)</sub> , (U <sub>13</sub> , U <sub>15</sub> , U <sub>3</sub> ) <sub>(0)</sub> , (U <sub>13</sub> , U <sub>15</sub> , U <sub>5</sub> ) <sub>(0)</sub>
고빈도 사용자 집합(L <sub>3</sub> )	(U <sub>1</sub> , U <sub>2</sub> , U <sub>3</sub> ) <sub>(2)</sub> , (U <sub>2</sub> , U <sub>3</sub> , U <sub>5</sub> ) <sub>(2)</sub> , (U <sub>2</sub> , U <sub>5</sub> , U <sub>15</sub> ) <sub>(2)</sub> , (U <sub>3</sub> , U <sub>13</sub> , U <sub>15</sub> ) <sub>(2)</sub>

Apriori 알고리즘을 여러 단계 이용하여 최종 고빈도 사용자 집합을 구성한다.

연관 사용자 집합에서 모든 연관 규칙과 신뢰도를 구한 후, ARHP 알고리즘을 이용하여 연관 사용자 군집을 한다. 여기서 <표 5>의 L<sub>3</sub>의 고빈도 사용자 집합에서 하이퍼 그래프의 분할을 위한 가중치는 연관 규칙의 평균 신뢰도를 사용한다. 결과적으로, 알고리즘에 의해 25명의

사용자를 <표 6>과 같이 3개의 연관 사용자 군집을 한다.

<표 6> 연관 사용자 군집

연관 사용자 군집	각 군집에 군집된 사용자
1	{U <sub>13</sub> , U <sub>14</sub> , U <sub>15</sub> , U <sub>16</sub> , U <sub>17</sub> , U <sub>18</sub> , U <sub>19</sub> , U <sub>20</sub> , U <sub>22</sub> , U <sub>23</sub> }
2	{U <sub>2</sub> , U <sub>3</sub> , U <sub>4</sub> , U <sub>5</sub> , U <sub>9</sub> , U <sub>10</sub> , U <sub>11</sub> }
3	{U <sub>1</sub> , U <sub>6</sub> , U <sub>7</sub> , U <sub>8</sub> , U <sub>12</sub> , U <sub>21</sub> , U <sub>24</sub> , U <sub>25</sub> }

연관 사용자 군집과 개인 선호도를 이용하여 사용자가 원하는 의상을 추천한다.

$$P_{u,j} = \frac{P_{set, item[j]}}{\sum_{i=1}^n P_{set, item[i]}} \times W_p + S_{set, item[j]} + \frac{P_{u, item[j]}}{\sum_{i=1}^n P_{u, item[i]}} \times W_p + S_{u, item[j]} \quad (\text{식 3-1})$$

- ( $P_{u,j}$  : 사용자  $u$ 의 콘텐츠  $j$ 에 대한 선호도,
- $P_{u, item[j]}$  : 사용자  $u$ 가 콘텐츠  $j$ 를 구매한 수,
- $n$  : 전체적인 아이템 수,
- $W_p$  : 콘텐츠 구매에 따른 가중치, 가중치에 따른 범위는 '0-1'사이의 피어슨 계수를 이용
- $S_{u, item[j]}$  : 사용자  $u$ 의 콘텐츠  $j$ 에 대한 검색율,
- $set$  : 사용자  $u$ 가 포함된 연관 사용자 군집)

위의 공식을 이용하여 추천된 여러 의상 중 데이터베이스의 콘텐츠 트랜잭션을 이용하여 사용자가 선택한 의상과 연관된 의상의 리스트를 작성한다. 작성된 리스트 의상은 아래의 식에 의해 각각의 선호도가 계산된다.

$$Total_j = P_{k,j} + \frac{R(a)_{u, set(a)[j]}}{\sum_{i=1}^n R(a)_{u, set(a)[i]}} \times W_{pa} + \frac{R(o)_{u, set(o)[j]}}{\sum_{i=1}^n R(o)_{u, set(o)[i]}} \times W_{po} \quad (\text{식 3-2})$$

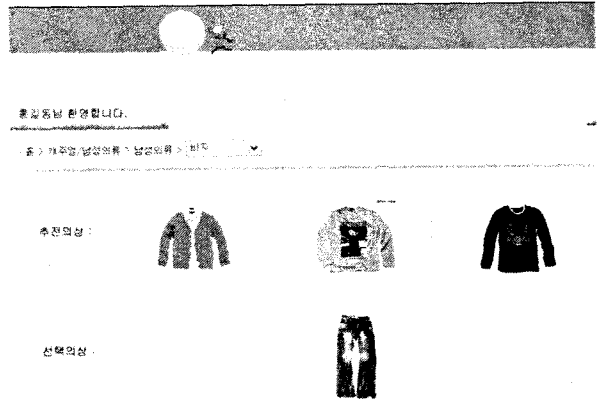
- ( $Total_j$  : 의상  $j$ 에 대한 전체 선호도,
- $k$  : 사용자  $u$ 가 선택한 의상,
- $j$  : 작성된 리스트 의상 중 하나,
- $P_{k,j}$  : 데이터베이스로부터 산출된 콘텐츠  $k$ 와  $j$ 의 연관율,
- $R(a)_{u, set(a)[j]}$  : 사용자  $u$ 가 포함된 연관 사용자 군집 중 사용자  $u$ 와 연관이 유사한 사용자  $set$ 이 기존에 구매했던 콘텐츠의 수,  $n$  : 콘텐츠의 수
- $W_{pa}$  : 연령에 부과되는 가중치
- $R(o)_{u, set(o)[j]}$  : 사용자  $u$ 가 포함된 연관 사용자 군집 중 사용자  $u$ 와 직업이 유사한 사용자  $set$ 이 기존에 구매했던 콘텐츠의 수
- $W_{po}$  : 직업에 부과되는 가중치)

계산된 의상의 선호도를 이용하여 높은 선호도를 가지는 의상들을 사용자에게 추천한다.

#### 4. 실험 및 평가

이 장에서는 우리가 제안한 의상 코디 시스템을 실험하기 위해서 사용자와 의상정보가 정확히 인식되고 의상 코디 서버에 전달되어 적절한 의상 추천 리스트를 제공하는지 평가한다. 실험 환경으로는 의상 코디의 구성 데이터베이스들은 시맨틱 웹에서 유기적으로 이용할 수 있도록 RDF 언어로 구성되어 있으며 웹 서비스는 AJAX로 구현하여 제공하였다. 데이터베이스는 연령층 별로 50명씩의 사용자들을 대상으로 설문 조사하여 구축하였다.

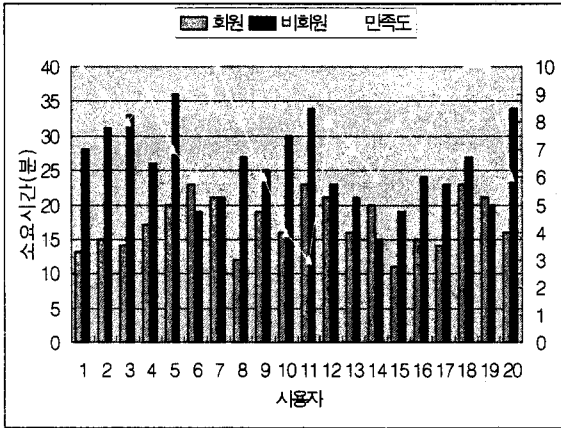
아이디가 홍길동이라는 사용자가 인터넷을 사용할 수 있는 장소에서 PC에 앉아 사이트에 접속을 한다. 사이트는 사용자가 ID와 비밀번호를 입력하면 사용자 정보 DB에서 방금 접속한 유저의 정보를 인식한다. 그리고 사용자의 입력 값에 의해 획득된 옷의정보를 가지고 의상 정보 DB에 저장되어 있던 옷의 정보와 위에서 제시한 (식 3-1)을 사용하여 적합한 의상을 찾은 후 의상 추천 서버에 보낸다. 서버에서는 사용자에게 적합한 의상을 추천해주고 사용자는 선택하게 된다. 사용자가 선택을 하게 되면 서버에서는 선택한 의상과 맞는 옷들을 (식 3-2)를 사용하여 사용자에게 추천해준다.



<그림 1> 의상 코디 웹서비스

위의 <그림 1>은 홍길동이라는 사용자가 Log-in한 후 검색창에 청바지라고 입력 하면 의상 추천 서버에 의해 사용자에게 적합한 의상 추천 리스트를 만들고 혼합 필터링을 사용하여 적합한 청바지를 추천해주면 사용자는 하나의 옷을 선택한다. 선택한 의상을 토대로 그림에서 보듯이 청바지와 가장 잘 어울리는 상의들을 다시 추천해준다. 실험 결과 위의 <그림 1>과 같이 사용자의 초기 의상 선택 후 사용자에게 가장 잘 어울리는 의상이 추천되었음을 알 수 있다.

참고 문헌



<그림 2> 추천 시스템 성능 측정 및 만족도

실험 사용자 집단의 테스트 결과 비회원 사용자에 비해 회원사용자의 의상선택 선택 소요시간이 약 33% 정도 감소하였고 또한 추천 시스템 사용 결과 사용자들의 만족도는 75.5%정도 만족했다는 결과가 도출되었다.

5. 결 론

본 논문에서는 추천의 효율성과 정확성을 증진시키기 위해 다중 혼합필터링을 이용한 시스템을 제안하였다. 의상을 검색시 데이터베이스에 저장된 낱말 관련 정보가 같은지에 대해 분석을 위해 내용기반 필터링을 사용하였다. 초기 추천의상의 선택 후 적절한 코드를 해주기 위해 사용자의 선호도와 의상 관련 정보들의 정확도를 높여줄 수 있는 협력적 필터링 및 연관 사용자 마이닝을 이용하여 이전보다 향상된 추천을 할 수 있었다. 기존의 필터링만을 통한 추천 모델이 지니고 있는 문제점을 해결하기 위한 방안으로, 시맨틱 웹을 이용하여 더 나은 만족도를 통해 추천 방법의 타당성을 입증하였다.

그러나 개인화 추천 시스템에서 주의할 점은 관련 정보들이 거짓이라면 결과가 잘못될 수 있다. 그러므로 수집된 정보가 거짓이 아님을 입증할 수 있는 방법에 대한 연구가 이루어져야 한다. 따라서 완전 자동화된 전자 프로파일링, 완벽히 개인화된 추천 시스템 구축은 앞으로 기술적으로 완성한다는 목표로 향후 관련 정보의 생성, 배포, 수집, 활용을 관리하는 시스템을 설계하여 예외적인 상황들을 고려하여 더욱 만족할만한 추천을 제공할 것이다.

[1] Luhn, G., Stoschek, B., Schilling, H., Kindelberger, A., Schurig, F., Vogel, M., Foerster, M., Kampfrath, R., Zschuppe, J., Adam, M., "Automation concept for complex production processes", Electronics Manufacturing Technology Symposium, Twenty-Fourth IEEE/CPMT, 1999.

[2] Sarwar, B. M., Karypis, G., Konstan, J. A. and Riedl, J., "Application of Dimensionality Reduction in Recommender System - A Case Study", ACM WebKDD 2000 : Web Mining for E-Commerce Workshop, 2000.

[3] Tjoa, A.M., Andjomshooa, A., Shayeganfar, F., Wagner, R., "Semantic Web challenges and new requirements", Database and Expert Systems Applications, 2005.

[4] Salter, J., Antonopoulos, N., "CinemaScreen recommender agent: combining collaborative and content-based filtering", IEEE Intelligent Systems and Their Applications, 2006.

[5] Amazon, <http://www.amazon.com/>, 2006.

[6] Lin, Lv., Yu-Shu, Liu., "Research of English Text Classification Methods Based on Semantic Meaning", Information and Communications Technology : ITI 3rd International Conference, 2005.

[7] Pollock, Stephen, "Rule-based message filtering system", ACM transactions on office information systems, 1988.

[8] Mooney, Raymond J., Roy, Loriene, "Content-based book recommending using learning for text categorization Proceedings of the ACM International Conference on Digital Libraries.", 2000.

[9] Good, N., Schafer, J.B., Konstan, J.A., Borchers, A., Sarwar, B., Herlocker, J., Riedl, J., "Combining collaborative filtering with personal agents for better recommendations", Proceedings of the AAAI- '99 Conference, 1999.