

클라이언트-서버 모델 기반의 개인화 텍스타일 감성 디자인 추천 시스템의 성능 평가

(Performance Evaluation of Personalized Textile Sensibility Design Recommendation System based on the Client-Server Model)

정 경 용 [†] 김 중 훈 ^{**} 나 영 주 ^{***} 이 정 현 ^{****}
(Kyung-Yong Jung) (Jong-Hun Kim) (Young-Joo Na) (Jung-Hyun Lee)

요 약 최근의 전자상거래 사이트들은 사용자 만족을 극대화 시키기 위해 사용자별로 개인화된 서비스를 제공하고 있다. 협력적 필터링은 개인화된 아이템 실시간 추천을 위하여 고안된 알고리즘이며 예측의 정확도와 성능을 증대시키기 위해서 다양한 보완 방법들이 제시되고 있다. 유용한 추천 시스템을 구축하기 위해서는 두 가지 문제를 동시에 고려해야 한다. 그러나, 협력적 필터링 기술에 관한 연구는 정확도 향상에 관한 연구가 주로 수행되었고 성능 문제는 간과하여 왔다. 본 연구에서는 추천 에이전트 시스템의 성능을 향상시킬 수 있는 대표 속성을 이용한 이웃 선택, 추천 텍스타일 집합, 유사 군집 요소를 협력적 필터링 기술에 추가하여 실험해 보았다. 패션 디자인 추천 에이전트 시스템(FDRAS)을 개발하여 시스템의 논리적 타당성과 유효성을 검증하기 위해 실험적인 적용을 시도하고자 한다.

키워드 : 협력적 필터링, 인지공학, 사용자 인터페이스

Abstract The latest E-commerce sites provide personalized services to maximize user satisfaction for Internet user. The collaborative filtering is an algorithm for personalized item real-time recommendation. Various supplementary methods are provided for improving the accuracy of prediction and performance. It is important to consider these two things simultaneously to implement a useful recommendation system. However, established studies on collaborative filtering technique deal only with the matter of accuracy improvement and overlook the matter of performance. This study considers representative attribute-neighborhood, recommendation textile set, and similarity grouping that are expected to improve performance to the recommendation agent system. Ultimately, this paper suggests empirical applications to verify the adequacy and the validity on this system with the development of Fashion Design Recommendation Agent System (FDRAS).

Key words : Collaborative Filtering, Cognitive Engineering, User Interface

1. 서 론

추천 시스템은 소수 혁신적인 전자상거래 사이트에서만 사용되어 왔으나 현재는 전자상거래를 새롭게 재구성하는 필수적인 비즈니스 도구가 되었다. 대형 웹사이

트들은 이미 추천 시스템을 사용하여 고객들의 구매활동을 지원하고 있다. 추천 시스템은 고객이 제공하는 다양한 유형의 정보로부터 학습하며 수많은 상품들 중에서 고객이 가장 좋아할 만한 것들을 추천해준다. 추천 시스템을 필터링 기법으로 분류해볼 때 규칙 기반 필터링, 학습 에이전트, 내용 기반 필터링, 협력적 필터링이 있다[1-5].

규칙 기반 필터링은 고객의 선호를 파악하여 정해진 규칙에 따라 상품을 특성과 연결시켜 추천을 제공한다. 이 규칙은 전문가에 의해 미리 정해지며, 이러한 방식은 상품 단가가 비싸거나 상품 자체의 특성이 복잡한 경우에 효과적이다. 내용 기반 필터링은 상품의 속성들에 대한 정보를 시스템이 보유하고 있다가 고객이 그 속성에 해당하는 키워드를 입력할 때 그 키워드에 해당하는 속

· 본 연구는 2003학년도 인하대학교의 지원에 의하여 연구되었음 (INHA-30220)

† 비 회 원 : 가천길대하 컴퓨터소프트웨어과 교수
kyjung@gcg.ac.kr

** 비 회 원 : 인하대학교 전자계산공학과
jhkim@nlsun.inha.ac.kr

*** 비 회 원 : 인하대학교 의류학과 교수
youngjoo@inha.ac.kr

**** 종신회원 : 인하대학교 컴퓨터공학부 교수
jhlee@inha.ac.kr

논문접수 : 2003년 11월 3일

심사완료 : 2004년 11월 17일

성을 지닌 상품들을 추천해주는 것으로, 정보검색 분야에서 많이 활용된다.

내용 기반 필터링은 아이템의 속성과 사용자의 흥미를 나타내는 사용자 프로파일을 비교하여 가장 동일한 아이템을 추천한다. 사용자 프로파일은 사용자가 평가한 아이템의 내용을 분석함으로써 생성시킬 수 있다. 아이템 속성은 아이템의 내용을 표현하는 특징과 아이템이 속하는 분류로써 표현된다. 이러한 내용 기반 필터링은 아이템의 특징이 다차원적으로 표현되므로 특징을 추출하는 전처리에 소요되는 시간이 상당히 길다는 단점을 갖는다. 이러한 다차원의 특징으로 인하여 내용 기반 필터링은 기계학습의 알고리즘에 적용하기가 쉽지 않다. 따라서 분류하기 위한 학습 방법에 따라 추천의 정확도가 달라진다는 단점을 갖는다.

협력적 필터링 개인화 기법은 고객들로부터 상품 선호도에 대한 데이터베이스를 구축하고, 새로운 고객이 등장하면 유사한 취향을 가진 이웃들을 데이터베이스로부터 찾아낸다. 그리고 이웃들이 선호하는 상품을 그 고객 역시 선호할 것이라는 가정 하에 상품을 고객에게 추천한다. 하지만, 연구분야나 실무분야에서 모두 매우 성공적으로 평가되어온 협력적 필터링 개인화 기법을 활용한 추천 시스템에 있어서 기본적으로 해결해야 할 두 가지 문제가 존재한다. 그 중에 하나는 협력적 필터링 개인화 알고리즘의 성능을 향상시키는 문제이다. 협력적 필터링 개인화 알고리즘은 실시간으로 수만 명의 잠재 이웃들을 탐색할 수 있지만, 실제 전자상거래 사이트에서는 수천만 명을 실시간으로 탐색할 수 있는 수준을 요구하고 있다. 또한 기존의 알고리즘은 사이트가 방대한 정보를 가지고 있을 경우 고객에게 추천을 제공함에 있어 심각한 성능 문제를 보여주고 있다. 두 번째 문제는 추천의 정확도를 향상시키는 것이다. 고객은 제품을 선택하는데 도움이 된다고 믿을 수 있는 추천을 원한다. 만일 고객이 추천 시스템을 신뢰한다면 추천된 제품을 구매할 것이고, 만일 그 제품이 마음에 들지 않으면 추천 시스템을 이용하지 않을 것이다[9,20]. 여기서 성능과 정확도 사이에는 서로 상충되는 점이 발생한다. 알고리즘이 이웃들을 검색하는데 소요되는 시간이 줄어들수록 성능이 향상되는 반면, 정확도가 떨어지기 때문이다. 이런 이유로 해서 유용한 해결책을 찾기 위해서는 성능과 정확도 문제를 동시에 고려하는 것이 중요하다[6,7].

2. 개인화 추천 시스템에 대한 연구

최근의 개인화 추천 시스템은 내용 기반 필터링과 협력적 필터링의 단점을 개선하기 위하여 이들 필터링을 병합하는 방법을 사용한다. 내용 기반 필터링과 협력적

필터링을 병합하는 기존의 연구는 여러 가지의 형태로 구분한다. 첫 번째의 병합 형태는 내용 기반과 협력적 필터링을 모두 단순히 병합하는 방법이다. 두 번째의 형태는 내용 기반 필터링에 협력적 필터링 방식을 적용하거나, 협력적 필터링 방식에 내용 기반 필터링을 적용하여 추천 목록을 생성하는 방법이다. 세 번째의 형태는 협력적 필터링 방법에 정보 필터링 기술을 이용한 개인화 에이전트를 병합하는 방법이다. 제안된 각각의 방법을 구체적으로 살펴보면 다음과 같다[21,22].

Cotter and Smyth는 내용 기반 필터링과 협력적 필터링의 방식에서 별도로 추천 목록을 생성하고, 마지막 결과에 이들 각각의 추천 목록을 단순히 병합하는 방법을 제안하였다. 목록의 병합은 각각의 방법에서 예측된 순위나 선호도를 평균함으로써 이루어진다.

Soboroff et al.는 LSI(Latent Semantic Indexing)을 통하여 특징-문서 행렬을 생성하고, 이 행렬의 값에 문서에 나타난 특징 빈도에 비례하여 주어진 가중치를 부여한다. 다음으로, 내용 프로파일 행렬을 생성하기 위하여 정의된 특징-문서 행렬과 협력적 필터링에 의한 선호도 평가 행렬을 곱하고, 이 행렬의 SVD를 계산한다. 마지막으로, LSI를 사용하여 내용 프로파일 행렬의 순위를 계산한다. 사용자를 표현하는 중앙값은 사용자에게 적합한 문서의 특징 벡터로부터 추출된다. 새로운 문서는 LSI 공간에서 각 사용자의 프로파일을 기준으로 순위가 부여된다. LSI를 이용한 필터링 방법은 입력 데이터의 차원을 축소하여 축소된 차원에서 추천할 수 있으므로, 상품 개수 혹은 고객 수가 많을 경우 드러나는 협력적 필터링의 문제점을 개선할 수 있다는 장점은 있으나 초기 평가 문제는 개선시키지는 못하였다는 단점을 갖는다.

Billsus와 Pazzani는 협력적 필터링의 희박성 문제를 개선하기 위하여 SVD(Singular Value Decomposition)를 사용한다. SVD에서는 사용자-아이템 행렬을 SVD에 의해 각기 다른 형식의 행렬로 분해한다. 선형적으로, 대부분의 독립적인 요소들은 매우 적으며, 더 작은 차원을 가지는 유사 모델을 만들 수 있다. 이러한 차원이 감소된 모델에서 사용자와 사용자, 아이템과 아이템, 사용자와 아이템의 유사성은 축소된 차원에서 발견될 수 있다. 이러한 방법에 따라 협력적 필터링의 희박성 문제를 개선할 수 있으며, 사용자에게 흥미로운 하나의 상품에 관련 상품까지 추천할 수 있다는 장점을 갖는다. 반면, 이 방법은 희박성 문제만을 개선하였을 뿐 초기 평가 문제를 개선하지 못하였다.

Basu et al.은 연역 논리 프로그램인 Ripper을 사용하여 사용자와 상품과의 관계를 학습하며, 사용자가 상품을 좋아할 것인가 아닌가를 예측한다. 예를 들어, 장

르 x의 영화를 좋아하는 사람과 사용자들이 좋아하는 장르 x와 같은 특성을 생성한다. 이 방법은 특징 추출을 “좋아하는가”와 “좋아하지 않는가”의 두 분류만으로 구분함으로써 다양한 형태의 추천에 적용할 수 없다는 단점을 갖는다. 또한, 이 방법은 초기 평가 문제를 개선하지 못하였다는 단점을 갖는다.

Pazzani는 역문헌빈도의 방법을 사용하여 훈련 집합으로부터 단어를 추출하고 가중치를 부여한다. 이 방법에서는 사용자의 프로파일을 가중치가 부여된 단어의 집합으로 표현하였다. 이 방법에서의 예측은 내용-프로파일의 행렬에 협력적 필터링을 적용함으로써 이루어진다. 여기서 내용-프로파일 행렬은 사용자들의 프로파일의 모임이다. 이 방법에서는 내용 기반 필터링에 협력적 필터링을 적용함으로써 협력적 필터링의 희박성 문제는 개선할 수 있었으나 초기 평가 문제는 개선하지 못하였다. 또한 특징 추출 방법으로 역문헌빈도를 사용하기 때문에 단어의 중의성으로 인하여 추천의 정확도를 저하시킨다는 문제점을 갖는다.

Balabanovic et al.의 Fab은 사용자의 연관 피드백과 “주제” 필터링을 통한 내용 기반 필터링을 하며, 이를 협력적 필터링과 병합하는 방법을 제안하였다. 내용 기반 필터링에서 문서는 주제 필터링에 의해 필터링되어 문서에 대한 순위 목록을 만들며, 생성된 목록에 대해 사용자는 연관 피드백을 제공함으로써 필터링이 이루어진다. 이 방법은 협력적 필터링의 초기 평가 문제를 개선할 수 있었으나 희박성은 개선하지 못하였다는 단점을 갖는다.

Good et al.은 여러 개인화 정보 필터링 에이전트와 협력적 필터링을 병합함으로써 아이템에 대한 추천을 제공한다. 새로운 사용자에 대한 예측은 새로운 사용자의 개인화 에이전트를 협력적 필터링 기술에 응용함으로써 생성된다. 이 방법은 협력적 필터링의 초기 평가 문제를 개선할 수 있었으나 희박성 문제를 개선하지 못하였으며, 또한 내용 기반 필터링의 다차원적인 특징 공간을 개선하지 못하였다.

3. 클라이언트-서버 모델 기반의 개인화 텍스타일 감성 디자인 추천 시스템

3.1 패션 디자인 추천 에이전트 시스템

감성 공학적 텍스타일 기반의 협력적 필터링 개인화 기술[5,8]을 이용하여 소재 설계하기 위한 패션 디자인 추천 에이전트 시스템은 고객의 감성을 파악하고 디자인 요소에 따른 감성 분석을 결합하는 추천 에이전트 시스템이다. 그림 1은 패션 디자인 추천 에이전트 시스템의 클라이언트 화면이다. 본 연구에서 설계한 패션 디자인 추천 에이전트 시스템을 FDRAS이라고 명명하였

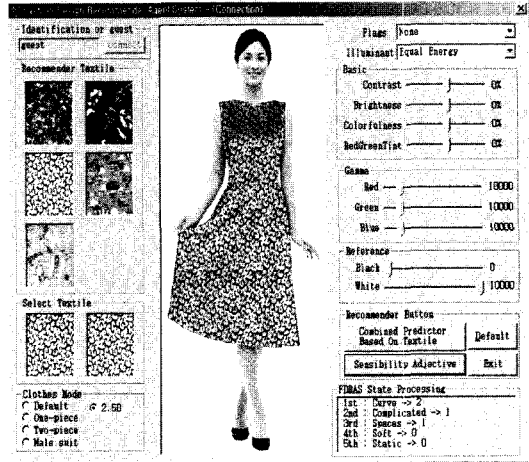


그림 1 패션 디자인 추천 에이전트 시스템(FDRAS)[9]

다[9].

FDRAS는 인하대학교 컴퓨터공학과와 의상학과의 공동 연구를 수행하였고, 본 연구의 패션 디자인 추천 에이전트 시스템은 프로그램 심의 조정 위원회에서 프로그램 저작권 등록[10]을 한 상태이다. 실제 구성도에 대한 작업 흐름도, 시스템 디자인에 대한 구체적인 세부 사항, 구현 부분에 대해서는 [9,10,11,14]의 선행 연구에서 내용을 언급하였다. 본 연구에서는 패션 디자인 추천 에이전트 시스템의 성능을 향상시킬 수 있는 방안에 대한 연구만을 기술한다.

본 추천 에이전트 시스템은 서버와 클라이언트 모듈로 구성하였다. 클라이언트 모듈을 설치할 수 있도록 요구하는 사용자들에게 무료 배포하였다. 네트워크 트래픽을 고려하여 클라이언트 프로그램에서 접속하면 서버에서는 데몬을 활성화하여 클라이언트의 요구한 질의에 대한 쿼리들을 분석하여 결과만을 전송한다. 전송 방식은 서버와 클라이언트간의 파일 전송 방식으로 본 시스템에 접속하는 모든 클라이언트에게 동시 전송 가능하게 멀티 에이전트 기반으로 구현하였다. 사용자의 성별과 나이를 얻기 위하여 주민등록번호 또는 guest를 이용하여 사용자 로그인을 할 수 있도록 하였다. 사용자 인증을 통하여 네트워크 외부로부터 허가 받지 않은 접속을 제한하였다. 이렇게 하여 네트워크 상에서 발생하는 예기치 못한 상황에 비교적 안정적으로 대응할 수 있다.

그림 2에서 패션 디자인 추천 에이전트 시스템의 핵심인 추천 엔진은 협력적 필터링 개인화 알고리즘, 유사 군집, 이웃 선정, 추천 텍스타일 집합, 연관 규칙 생성

1) 프로그램심의조정위원회 프로그램 저작권 등록(등록번호 2004-01-173-003676)

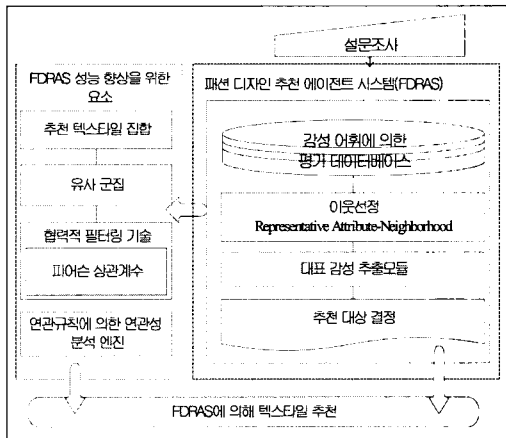


그림 2 패션 디자인 추천 에이전트 시스템 구성도

엔진의 다섯 부분으로 구성되어 있다. 협력적 필터링 개인화 기법은 고객의 선호에 대한 감성 데이터베이스로부터 고객들간의 유사성을 분석한 후 텍스타일에 대한 선호도를 예측해주는 역할을 한다. 고객의 성향을 바탕으로 비슷한 취향을 가진 고객들을 유사 군집으로 묶음으로써 집약된 선호 정보를 협력적 필터링 개인화 기법에 활용할 수 있게 한다. 유사 군집 분석에 포함된 선호 정보는 군집의 중심 값으로 대표된다. 만일 유사 군집 분석에서 제외된 고객이나 신규 고객의 경우라면 기존의 군집 중심 값에 협력적 필터링 개인화 기법을 적용한다. 따라서 군집을 하나의 고객으로 간주하면 된다. 고객 정보를 사용하지 않고 유사 군집 분석을 통한 집약된 선호 정보를 사용함으로써 정확도를 다소 희생시키더라도 성능을 향상시키고자 하는 것이다.

웹을 이용한 설문을 통해서 사용자의 감성이나 선호도에 대한 텍스타일의 대표 감성어휘를 추출함으로써 소개 개발을 위한 감성 어휘 데이터베이스를 구축한다. 여기서 어떤 텍스타일에 대해 고객이 어떠한 감성을 느끼는지를 알면 되나 그 감성을 직접 측정하는 것이 어렵다. 따라서 감성 표현에 가장 가까운 어휘를 통해 간접적으로 측정하는 수단을 사용하였다. 수집된 감성어휘 데이터베이스는 쌍극 형용사 형태로 재구성하며 통상 5 단계로 구성하여 인간이 가지는 감성을 조사하고, 요인 분석으로 데이터를 분석하였다.

협력적 필터링 개인화 기법을 이용하여 텍스타일 추천에 사용될 이웃의 수를 결정하기 위해서 Representative Attribute-Neighborhood 방법을 사용한다 [9,13-15]. 사용자들간의 유사도 가중치를 계산하기 위해서 피어슨 상관계수를 사용한다. 유사도 계산을 통해 얻은 유사 사용자들은 통하여 텍스타일에 대한 선호도를 예측하는 방법으로는 유사 사용자의 선호도 가중치

합을 사용한다[11,12]. 구축된 감성어휘 데이터베이스를 기반으로 성향이 비슷한 사용자에게 텍스타일을 추천한다[10]. 즉, 사용자는 패션 디자인 추천 에이전트 시스템의 클라이언트에서 선호감성으로서 「선호감성어휘 추천」을 통해서 감성어휘를 입력하면 서버에서 감성이 비슷한 사용자들을 기반으로 사용자를 군집한 후 협력적 필터링 개인화 기법[5,8,21]에 의해 검색된 텍스타일 5종류를 서버에서 전송하여 추천하게 된다.

3.1.1 연관성 분석을 이용한 추천 텍스타일 집합

Apriori 알고리즘[15,21]을 이용한 연관성 분석은 아이템간의 관계를 분석하기 위해서 사용되지만 연관성 분석을 사용자간의 연관 관계로 확장 적용하여 개인화된 텍스타일 디자인 추천을 위해 사용할 수 있다. 선호도를 연관성 분석에 맞게 매핑하고 사용자 연관성(user association)과 텍스타일 디자인 연관성(textile design association)을 분석해 사용자에게 개인화된 텍스타일 디자인을 추천할 수 있다.

3.1.1.1 연관성 분석의 정의

연관성 분석에 대한 기본적인 정의를 하면 트랜잭션은 텍스타일에 대한 집합이고, 연관성 규칙은 “A→B”로 표현되어지는데 A와 B는 텍스타일이다. 연관성 규칙은 하나의 트랜잭션 안에 존재하는 텍스타일A는 같은 트랜잭션 안에서 동일한 확률을 가지고 텍스타일B를 포함한다는 것을 의미한다. 각각의 연관성 규칙은 주어진 트랜잭션의 집합과 관련해서 신뢰도(Confidence)와 지지도(Support) 2개의 기준을 가진다. 신뢰도는 텍스타일A를 포함하고 있는 트랜잭션들 중 텍스타일B를 포함하고 있는 트랜잭션의 비율이고 신뢰도를 결정하기 위한 식 (1)은 다음과 같이 정의한다.

$$\text{신뢰도}(A \rightarrow B) = \Pr(A|B) \quad (1)$$

지지도는 분석 대상이 되는 전체 트랜잭션들 중 A와 B를 모두 포함하고 있는 트랜잭션의 비율이다. 지지도를 결정하기 위한 식 (2)는 전체 텍스타일들의 쌍 중에 각 연관 텍스타일의 출현 빈도를 나타낸다. 식 (2)는 텍스타일 A와 B의 모든 항목을 포함하고 있는 트랜잭션의 수를 데이터베이스 내의 전체 트랜잭션의 수로 나눈 결과 값을 나타낸다.

$$\text{지지도}(A \rightarrow B) = \Pr(A \cup B) \quad (2)$$

다른 용어로 표현을 하면 신뢰도는 텍스타일 집합 사이의 관계(Correlation)의 정도이고, 지지도는 텍스타일 집합 사이의 관계의 중요성(significance)로 표현될 수 있다. 결국 연관성 분석은 주어진 트랜잭션에서 사용자가 정의한 최소 지지도와 최소 신뢰도 보다 큰 모든 연관성 규칙을 발견하는 것이다.

3.1.1.2 Apriori 알고리즘

Apriori 알고리즘을 이용해서 연관성 규칙을 발견하기

위해서 먼저 최소 지지도 이상을 만족하는 트랜잭션으로 구성된 아이템 집합을 계산한다. 아이템 집합의 지지도는 아이템 집합을 포함하고 있는 트랜잭션의 수이고 최소 지지도를 가진 아이템 집합을 Large 아이템 집합이라 한다. 다음 단계에서는 전단계에서 Large 아이템 집합으로 발견된 아이템 집합을 기반으로 아이템 집합의 아이템 개수가 증가된 새로운 Large 아이템 집합을 구하고, 이를 후보(candidate) 아이템 집합이라 부른다. 후보 아이템 집합의 지지도를 계산하여 지지도 이상인 집합을 다시 다음 단계의 Large 아이템 집합으로 간주한다. 이러한 단계는 새로운 Large 아이템 집합이 발견되지 않을때 까지 계속된다.

Apriori 알고리즘은 후보 아이템 집합을 생성할 때 데이터베이스의 모든 트랜잭션을 검색하는 것이 아니라 직전 단계의 Large 아이템 집합만을 사용한다. 직관적으로는 Large 아이템 집합의 부분집합은 Large 아이템 집합이 되는 것이다. 그래서 k개의 아이템을 가지고 있는 후보 아이템 집합은 k-1개의 아이템을 가진 Large 아이템 집합을 조합해서 생성되고 Large가 아닌 부분집합을 가진 것들은 삭제된다. 이 과정으로 후보 아이템 집합을 매우 적게 생성할 수 있다. Apriori 알고리즘은 다음과 같다.

```

L1 = {Large 1-itemsets}; //빈말 항목을 구성
For (k=2; Lk-1 ≠ ∅; k++) do begin
    Ck = apriori-gen(Lk-1); //새로운 후보항목
    Forall transactions t∈D do begin
        Ct = subset(Ck, t); //t에 포함된 후보항목
        Forall candidates c∈Ct do
            c.count++;
    End
    Lk = {c∈Ck | c.count ≥ minsup}
    //최소 지지도 이상의 항목의 조합을 추출
End
Answer = ∪k Lk;
    
```

k-itemset은 k개의 아이템을 가진 집합이고, L_k는 최소 지지도를 가진 Large k-itemsets의 집합이다. L_k의 구성원들은 아이템 집합과 지지도 count 2개의 필드를 가진다. 첫번째 과정에서 단순히 최소 지지도를 넘는 1개의 아이템으로 구성된 Large 1-itemsets을 결정한다. 이후 2-itemset, 3-itemset, ..., k-itemset으로 확장이 되는데 이 연속적인 과정에서 k-itemsets을 만드는 과정을 보면 후보 아이템 집합인 C_k를 계산하기 위해서 k-1번째 과정에서 발견되는 Large itemsets인 L_{k-1}을 사용한다. 후보 아이템 집합인 C_k를 구하는 함수를 apriori-gen으로 정의하며 C_k = apriori-gen(L_{k-1})으로 표현한다. 후보 아이템 집합인 C_k를 생성한 후 C_t의 후보들이 최소 지지도를 만족하는지를 주어진 트랜잭션 t

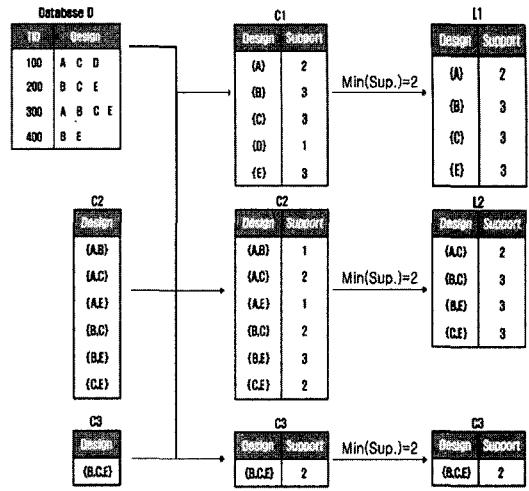


그림 3 연관 규칙을 유추하는 과정의 예

에서 점검하는데 이 과정은 C_t = subset(C_k, t)으로 표현된다. 그리고 마지막으로 최소 지지도를 점검하여 최소 지지도를 넘는 후보들로 Large itemsets인 L_k를 생성한다. 그림 3은 Apriori 알고리즘을 이용하여 연관 규칙을 유추하는 과정이다.

3.1.1.3 Apriori 후보 집합 생성

Apriori 알고리즘에서 후보 아이템 집합을 만드는 함수인 apriori-gen 함수는 모든 후보들로 Large (k-1)-itemsets의 집합인 L_{k-1}을 인수로 한다. 이 apriori-gen 함수는 조인 단계와 삭제 단계가 있다. 조인 단계는 다음과 같다.

```

Insert into Ck
Select p.item1, p.item2, ..., p.itemk-1, q.itemk-1
From Lk-1.p, Lk-1.q
Where p.item1= q.item1, ..., p.itemk-2= q.itemk-2, p.itemk-1
< q.itemk-1
    
```

Large (k-1) itemsets을 조인해서 k개의 아이템으로 구성된 후보 아이템 집합 C_t를 만드는데 p로 정의된 L_{k-1}과 q로 정의된 L_{k-1}에서 사전순서대로 정의된 아이템들의 최소 k-2개의 공통된 아이템을 가진 아이템 집합을 찾고 서로 다른 k-1번째 아이템을 순서대로 나열하여 새로운 k개로 구성된 후보 아이템 집합을 만든다. 조인 단계는 다음과 같다.

```

Forall itemsets c∈Ct do
    Forall (k-1)-subset s of c do
        If (s ≠ Lk-1) then
            Delete c from Ct;
    
```

조인 단계에서 만들어진 후보 아이템 집합 C_k의 구성원인 c중에서 c의 (k-1)-subset이 L_{k-1}에 속하지 않는

것을 찾아서 후보 아이템 집합에서 삭제한다. 조인 단계와 삭제 단계를 예를 들어 설명하면 L_3 가 $\{(1,2,3), (1,2,4), (1,3,4), (1,3,5), (2,3,4)\}$ 일 때 조인 단계에서 $\{(1,2,3)$ 과 $\{(1,2,4)$ 에 의해서 $\{(1,2,3,4)$ 가 생성되고, $\{(1,3,4)$ 와 $\{(1,3,5)$ 에 의해서 $\{(1,3,4,5)$ 가 생성되어 C_4 는 $\{(1,2,3,4), (1,3,4,5)\}$ 가 된다. 삭제 단계에서는 $\{(1,3,4)$ 와 $\{(1,4,5)$ 가 L_3 에 있는지를 검사하게 되는데 $\{(1,4,5)$ 가 존재하지 않으므로 $\{(1,3,4,5)$ 를 삭제하고 C_4 에는 $\{(1,2,3,4)$ 만 남게 된다.

3.1.1.4 연관성 분석을 위한 데이터 매핑

연관성 분석은 구매/비구매, 혹은 선호/비선호와 같이 바이너리한 데이터 정의가 필요하다. 즉, 같이 구매되거나 같이 선호되는 아이템의 집합을 찾아내는 것이므로 선호도 데이터를 가지고 있을 경우 평균적인 선호도나 기준 선호도 값을 정해 기준값 이상인 선호도는 선호로 정하고 기준값 이하인 경우에는 비선호로 값을 이분화한다. 선호도 데이터로 구매나 입찰 등의 암목적 데이터를 이용했을 경우 아이템에 대한 구매여부나 입찰여부 등을 연관성 분석을 위한 데이터로 사용하면 된다. 위의 데이터를 기반으로 사용자 연관성과 텍스트일의 연관성을 분석한다[15]. 사용자의 연관성은 식 (3)과 같이 표현한다.

$$\begin{aligned} &(\text{사용자A, 사용자B}) \rightarrow (\text{사용자C}) & (3) \\ &\text{신뢰도 90\%, 지지도 30\%} \end{aligned}$$

식 (3)은 사용자A와 사용자B가 선호하는 텍스트일 디자인의 90%는 사용자C가 선호하며 사용자A, B, C는 모든 아이템 중 30%를 공통으로 선호한다는 것을 의미한다. 결국 텍스트일 디자인을 중심으로 선호한 사용자들의 연관성 분석하는 것이다. 텍스트일 연관성은 식 (4)와 같이 표현된다.

$$\begin{aligned} &(\text{텍스트일A, 텍스트일B}) \rightarrow (\text{텍스트일C}) & (4) \\ &\text{신뢰도 90\%, 지지도 40\%} \end{aligned}$$

식 (4)는 텍스트일A와 텍스트일B가 포함된 트랜잭션 중 90%가 텍스트일C를 포함하고 텍스트일A, B, C는 모든 텍스트일 집합 중 40%에 공통으로 포함된다는 것을 의미한다. 결국 사용자를 중심으로 텍스트일의 연관성을 분석하는 연관성 분석의 전형적인 방법[1,15,20]이다.

3.1.1.5 연관성 분석을 통한 텍스트일 디자인 추천

실험을 위해 테스트 집합과 훈련 집합으로 데이터를 분리하고 테스트 집합에 있는 사용자를 타겟 사용자라, 훈련 집합에 있는 텍스트일을 타겟 텍스트일로 정의한다. 먼저, 사용자의 연관성을 이용해서 추천하는 방법에 대해서 설명하면, 사용자 연관성을 통해서 식 (4)와 같은 규칙이 생성되는데 여기서 사용자A와 사용자B는 훈련 사용자이고, 사용자C는 타겟 사용자라 할 때 사용자

C의 타겟 텍스트일에 대한 스코어를 예측하기 위해서는 사용자A와 사용자B가 타겟 텍스트일을 선호했다는 전제가 필요하며 만약 사용자A와 사용자B가 타겟 텍스트일을 선호했다면 사용자C의 타겟 텍스트일에 대한 예측된 선호도는 지지도와 신뢰도의 곱으로 표시된다. 사용자A, B, C를 포함하고 있는 모든 규칙에 대해서 타겟 텍스트일에 대한 예측된 선호도를 계산해서 합계를 합으로써 각각의 타겟 텍스트일에 대한 선호도를 계산한다. 계산되어진 선호도의 순으로 타겟 텍스트일을 추천하게 된다.

텍스트일의 연관성을 통해서 식 (4)와 같은 규칙이 생성되는데 여기서 텍스트일A와 텍스트일B는 훈련 텍스트일이고, 텍스트일C는 타겟 텍스트일이 된다. 타겟 사용자가 훈련 텍스트일인 텍스트일A와 텍스트일B를 선호했을 때 타겟 텍스트일인 텍스트일C에 대한 예측된 선호도는 텍스트일A, B, C를 포함하고 있는 규칙의 지지도로 정의되고 이러한 지지도의 순으로 타겟 텍스트일을 추천하게 된다.

이와 같이 연관성 분석을 이용해서 텍스트일을 추천하는 방식은 의미있는 결과를 보이지만 실시간 추천을 위해서는 적절하지 못하다. 실시간 추천의 의미 중 사용자의 반응에 즉각적으로 추천을 해줄 수는 있지만 사용자의 행동을 반영한 추천은 연관성 분석이 오래 걸리는 작업인 점을 고려하면 불가능하므로 이를 보완하여 적용하는 방법이 필요하다. 본 연구에서는 연관성 분석을 통한 추천 텍스트일 집합을 도입하였다. 연관성 규칙을 협력적 필터링 개인화 기법에서 사용자들간의 유사도 가중치를 계산하는 부분에 적용하는데 두 사용자간에 공통으로 선호도를 보인 텍스트일의 집합이 연관성 분석의 규칙들 안에 포함되어 있으면 두 사용자는 많은 다른 사용자에게 의해서 공통적으로 선호된 텍스트일을 선호하는 것이므로 두 사용자간 유사도 가중치를 높여 주어야 한다는 것이 연관성 분석을 통한 추천 텍스트일 집합의 아이디어이다. 연관성 분석을 통해서 유사도 가중치에 보완할 가중치는 식 (5)와 같다.

$$A(a, i) = \text{Normalization} \left(\frac{\sum_{v_1, \dots, v_n \in I_{a,i}} \text{Support}(v_1, \dots, v_n)}{\text{TotalUserCount}} \right) \quad (5)$$

$A(a,i)$ 는 사용자 a와 사용자 i의 n개의 텍스트일을 가진 연관성 규칙 벡터를 통해 계산된 연관성 분석 가중치이다. $\text{Support}(v_1, \dots, v_n)$ 은 연관성 규칙 벡터 v_1, \dots, v_n 의 지지도이고, $I_{a,i}$ 는 사용자 a와 사용자 i가 공통으로 선호도를 평가한 텍스트일 집합을 포함하고 있는 연관성 규칙 벡터들의 지지도의 합계한 후 전체 사용자 수로 나눈 값을 0과 1사이의 값으로 정규화 한다. 두 사용자의 유사도 가중치 값에 식 (5)에서 계산한 연관성 분

석 가중치를 증가시킴으로써 연관성 규칙 안에 포함되어 있는 텍스타일을 공통으로 선호하고 있는 사용자들을 협력적 필터링 개인화 기법의 텍스타일 선호도 예측에서 더 강조하는 것이다. 이를 위해서 적절한 최소 지지도 값에 따라 생성되는 연관성 규칙 벡터의 수와 지지도의 합계 값의 적정수준을 정해야 한다.

3.2 본 연구에서 사용한 알고리즘

3.2.1 기존 연구에서 채택한 알고리즘

본 연구에서 설계된 패션 디자인 추천 에이전트 시스템은 기존 선행 연구의 성과를 토대로 하여 추천 엔진을 설계하였다. 기존 연구에는 피어슨 상관계수를 기본 알고리즘으로 하였다.

기본적으로 텍스타일 기반의 협력적 필터링 개인화 기법[3,4,-15]은 4가지 단계를 거쳐 구현, 평가되는데 그 단계는 표 2와 같다.

표 2 협력적 필터링 개인화 기법의 단계

단계1	활성화된 사용자와 이웃들과의 유사도 가중치를 정의한다.
단계2	활성화된 사용자의 특정 텍스타일을 예측하기 위해서 유사도가 높은 이웃을 어떤 기준으로 선택할지를 결정한다.
단계3	유사한 선호도를 가지는 이웃들의 텍스타일에 대한 감성 선호도를 기반으로 활성화된 사용자의 텍스타일을 예측한다.
단계4	실제적인 텍스타일의 선호도와 예측값을 적절한 기준으로 평가한다.

기존의 추천 시스템에 대한 연구는 정확도를 향상시키는 알고리즘에 대한 연구가 진행되었다. 따라서 패션 디자인 추천 에이전트 시스템의 엔진에도 피어슨 상관계수 방식을 채택하였으며, 정확도 향상에 기여한다고 평가된 요소들을 추가하였다. 표 3은 Herlocker의 연구 [8]로부터 채택한 알고리즘 요소들을 나타낸다.

표 3 Herlocker의 연구로부터 채택한 요소

요소	적용방식
Significant 가중치	
이웃들간의 상관계수 계산의 신뢰도 문제를 반영	n/50
이웃 선택	
예측을 계산하기 위해 사용자의 집합을 선별하는 방식	k명의 가장 좋은 이웃 선택
평가 정규화	
예측 계산할 때 정규화를 함으로써 정확도를 향상시키는 방식	Bias form mean

3.2.2 성능 향상을 위해 추가된 요소

패션 디자인 추천 에이전트 시스템의 성능을 향상시

키고자 이웃 선택, 추천 텍스타일 집합, 유사 군집을 추가 하였다. 사용자간의 유사도 가중치를 계산하고 특정한 사용자의 패션 디자인에 대한 선호도를 예측하기 위해서는 이웃으로 사용할 사용자 수를 적절하게 결정해야 한다. 유사도 가중치가 구해진 모든 이웃들을 사용해서 선호도를 예측할 수 있으나, 이러한 방법은 정확도나 성능면에서 권장할 사항은 아니다. 반면 너무 높은 유사도의 이웃들만을 예측에 사용할 경우에는 다른 고객들과 유사도가 높지않은 사용자의 텍스타일에 대해서는 예측할 수 없는 문제점이 발생한다. 그러므로 시스템이 예측할 수 있는 적절한 이웃의 수를 결정하는 것이 중요하다. 따라서 예측에 사용될 이웃의 수를 결정하기 위해서는 최적의 임계값과 최적이웃을 선정해야 한다. 임계값은 사용자간의 유사도 가중치가 기준값 이상인 이웃들만을 사용해서 예측하도록 하는 방법이고, 최적 이웃은 특정 사용자와 유사한 n명의 이웃을 사용해서 예측하도록 하는 방법이다. 위의 두 방법을 조합하여 유사도 가중치가 어느 값 이상인 이웃들 중 n명을 사용할 수 있도록 하는 것이 바람직한 방법이다. 이웃 선택은 계산 시간의 단축뿐만 아니라 유사 사용자만을 계산에 사용하기 때문에 잡음을 줄여들어 예측 값의 정확도가 개선된다. 하지만 일정수준 이하의 인원이 이웃으로 사용되어진다면 충분하지 못한 양의 자료가 선호도 예측에 사용되어져서 올바르지 못한 결과를 나타내기도 한다. 또한 이웃 선택은 유효 범위를 낮추는 작용을 하기도 한다. 즉, 연산에 참여하는 사용자의 수를 줄임으로써 현 사용자와 동일한 텍스타일을 평가한 사용자의 수를 줄여 선호도를 예측할 수 없는 경우를 증가시킨다. 따라서 선호도 예측에서 이웃의 선택 기준은 중요한 요소라고 할 수 있다.

본 연구에서는 위의 조합방법으로서 성별과 나이에 의한 Representative Attribute - Neighborhood 방법을 사용하였다. 본 연구에서는 대표 속성에 의한 이웃 선택이 성능을 향상시킬 수 있다는 가정 하에 성능 향상을 위한 요소로 이웃 선택을 추가하였다. 성별 또는 같은 나이를 가진 사람들은 각각의 텍스타일에 대해서 유사한 선호도를 가진다고 가정한다. 즉, 남성과 여성간의 성별 차이와 세대차를 통해서 예측의 정확도를 높이기 위함이다.

성능 관점에서 보면 협력적 필터링 개인화 기법을 이용하여 모든 텍스타일에 대해 예측 값을 계산하고, 예측 값이 가장 높은 텍스타일을 추천하는 기존의 방식은 비효율적이다. 텍스타일에 대한 예측을 계산하는데 수많은 질의와 계산이 데이터베이스에서 수행된다. 이러한 방식은 데이터베이스의 규모가 커질수록, 즉 사용자와 텍스타일의 수가 증가할수록 질의 횟수와 계산 횟수가 지수

적으로 증가하게 되어 시스템의 과부하를 야기하게 된다. 결과적으로 추천의 반응시간은 증가하게 된다. 이러한 문제점을 해결하기 위해서 본 연구에서는 추천 텍스트타일 집합이라는 개념을 도입하였다. 추천 텍스트타일 집합이란 사용자에게 몇 가지 텍스트타일만을 추천한다고 할 때, 평가하지 않은 모든 텍스트타일에 대해 예측할 것이 아니라 사용자가 좋아할 만한 텍스트타일의 집합을 의류 소재 전문가에 의해 추출함으로써 질의 횟수와 계산 횟수를 줄이는 방법이다. 특정 사용자의 추천을 위해 Representative Attribute - Neighborhood 방법에 의한 이웃 선택에서 일정 점수 이상을 부여한 텍스트타일 중 사용자가 평가하지 않은 텍스트타일을 추천 텍스트타일 집합으로 정하였다.

추천 텍스트타일 집합이 텍스트타일의 측면에서 사이즈를 줄이는 것이라면, 유사 군집은 사용자 측면에서 사이즈를 줄이는 방식이다. 즉 유사한 취향을 가진 사용자들간의 군집하고 예측할 때 선호도를 이용하는 것이 아니라 군집의 중심 값을 이용함으로써 계산 시간을 단축시킬 수 있다. 그리고 사용자들의 수 증가에 관계없이 일정한 반응시간을 유지할 수 있을 것이라는 가정 하에 유사 군집을 본 연구의 요소에 포함시켰다. 사용자 별로 텍스트타일에 대해 평가한 경우에는 평가 값을 사용하였고, 결측값이 존재하는 경우에는 협력적 필터링 개인화 기법에 의해 예측 값을 적용하였다. 데이터베이스의 희박성 문제로 인하여 예측을 계산할 수 없을 경우 유사 군집 분석에서 제외시켰다.

3.3 평가 데이터

성능 평가를 위한 데이터는 설문 조사를 통하여 평가 데이터를 구성하였다. 텍스트타일에 대한 감성을 평가하기 위한 감성어휘는 사전, 잡지, 문헌 등에서 단어(감성어휘)를 추출하는 방법을 사용하여 18개의 감성어휘 쌍을 확정하였다. 18개의 감성어휘는 어두운, 밝은, 시골스런, 도시적인, 청순한, 색시한, 보수적인, 개방적인, 탁한, 맑은, 곡선적인, 직선적인, 정적인, 동적인, 복고적인, 현대적인, 차가운, 따뜻한, 단조로운, 복잡한, 부드러운, 거친, 동양적인, 서양적인, 어린이다움, 어른스런, 품위없는, 품위있는, 기계적인, 자연적인, 여성적인, 남성적인, 낡은듯한, 새로운, 성긴, 치밀한이다. 의미분별척도의 형식으로 -2에서 +2까지의 척도(5단계)로 942명의 사용자(남자 512명, 여자 430명)들에게 평가하였다. 웹에서 설문 받도록 ASP와 Microsoft SQL Server 2000을 사용하여 프로그래밍하였다. 각 알고리즘은 Visual Studio C++ 6.0을 사용하여 구현하고 시뮬레이션 하였다. 감성 데이터베이스를 구축하기 위한 DBMS는 Microsoft SQL Server 2000을 활용하였고, 감성어휘 및 사용자가 평가한 선호도의 요인 분석을 위해 활용되는 소프트웨어로

UNIX용 네트워크 버전의 SAS 패키지를 사용하였다. 설문한 사용자들은 컴퓨터 모니터를 통하여 텍스트타일을 보면서 준비된 감성어휘로 평가하도록 하였다. 컴퓨터 모니터를 통하였으므로 색상의 동질성이 충분하지 못하다는 점이 본 연구의 제한점이라 볼 수 있다. 이렇게 수집된 평가 데이터를 기반으로 텍스트타일 데이터베이스를 구축하였다. 텍스트타일 데이터베이스는 사용자가 평가한 평가 데이터, 사용자 프로파일, 텍스트타일에 대한 정보로 구축하였다. 설문조사는 인하대학교 컴퓨터.정보공학과 HCI 연구실 서버에서 진행하였다. 설문조사 URL은 <http://hci.inha.ac.kr/sulTDS> 이다. 설문조사에 대한 결과 및 특정 텍스트타일에 대한 대표 감성어휘 추출의 전체 결과 화면은 <http://HCI.inha.ac.kr/sulTDS/result.htm>에 나타내었다. 그림 4는 설문조사의 결과로 웹에서의 감성어휘별 평가 데이터의 분포이다.

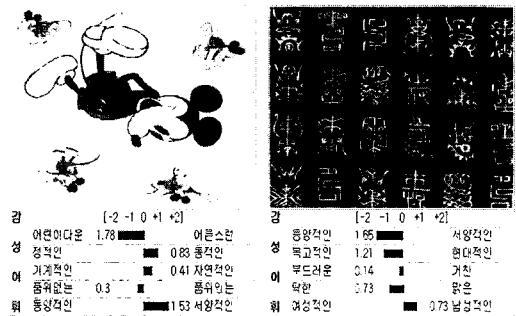


그림 4 감성어휘별 평가 데이터 분포[9]

텍스트타일은 상상물, 자연물, 기하학물, 인공물의 4개의 카테고리로 각각 1,240개의 텍스트타일을 선별하였다. 디자인 요소 추출과 결정은 선행 연구[1,9,11,12,17,19]를 참조하여 추출하였으며, 무작위 표집에 의해 모티브의 소재를 중심으로 분류하고 그 분류 속에서 디자인 요소가 가능한 골고루 포함되도록 배려하여 최종 1,240개가 선별되었다. 디자인 요소별로 고른 분포를 지니도록 표본의 침착을 거듭하였으나 존재하지 않는 조합도 있으므로 모든 디자인 요소들의 분포가 완벽하게 고르다고는 볼 수 없다[18]. 텍스트타일은 무작위로 배열하여 구성되도록 설문을 진행하였다. 또한 1,240개 텍스트타일에 18쌍의 감성어휘 중에서 무작위로 5쌍을 선택하여 배정하였다. 디자인 요소로는 모티브대 배경의 비율(대, 중, 소), 모티브의 변화도(대, 중, 소), 해석법(Realistic, Stylized, Abstract, Geometric), 모티브의 배열(all-over, 1-way, 2-way, 4-way), 모티브의 명료성(대, 중, 소), 채도차(대, 중, 소) 등이 되며 이는 의류 소재 전문가에 의해서 평가[16-19]되었다.

4. 성능 평가

4.1 기존 연구의 평가 기준

기존의 연구에서 추천 알고리즘을 평가하는 기준은 두 가지가 있다. 시스템의 적용범위가 얼마나 되는가 하는 측면과 알고리즘의 정확도 측면으로 나누어 볼 수 있다[20].

4.1.1 시스템의 적용범위(Coverage)

Coverage는 추천 시스템이 예측 값을 제공할 수 있는 전체 아이템의 비율로써 계산된다. 시스템에서 이웃 사람들의 크기나 이웃을 발견하기 위한 사용자의 수가 작으면 일반적으로 Coverage는 감소된다. Herlocker의 연구[3,8]에서 Coverage는 모든 사용자들이 예측 값을 요구한 텍스트일 수와 시스템이 예측 값을 제공할 수 있었던 텍스트일의 수의 비율로써 결정되었다.

Coverage는 예측의 대상이 되는 (사용자, 텍스트일) 벡터 집합의 벡터 개수를 전체로 보았을 때 특정한 알고리즘이 예측해 내는 개수의 전체 대비 비율이다. 예측에 사용되는 이웃의 수가 적으면 Coverage는 줄어들게 된다. 또한 새로운 텍스트일과 같이 전체 사용자중 어떤 사람에 의해서도 평가되지 않은 텍스트일이 예측의 대상이 되면 이 텍스트일의 경우에는 예측이 불가능할 것이며, 소수의 사용자에 의해서 선호도가 평가된 텍스트일의 경우에도 사용자간 유사도가 0이 될 가능성이 존재하므로 선호도를 예측할 수 있다. Coverage는 전적으로 협력적 필터링 개인화 기법의 정확도와 성능과 반비례하는 경우가 많고 적절한 집점을 찾는 것이 중요하다.

4.1.2 정확도

협력적 필터링의 정확도를 측정하는 방법으로 여러 가지 지표가 있는데, 통계적인 정확도와 의사결정 정확도라는 두 가지 범주로 나누어 볼 수 있다. 통계적인 정확도는 협력적 필터링 개인화 기법에서 산출된 예측 값이 실제 사용자가 평가한 수치와 얼마나 차이가 있는지를 평가하는 기준으로 보통 Mean Absolute Error (MAE)가 사용된다. 통계적 정확도의 평가 기준으로 MAE는 대부분의 기존 연구에서 광범위하게 사용되었다. Sarwar와 Herlocker의 연구에서 모두 MAE를 예측 정확도 측정 기준으로 사용[6-9,20-22]하였다.

4.2 본 연구의 평가 기준

본 연구에서는 평가 기준으로는 크게 예측의 정확도에 대한 평가기준과 알고리즘의 성능에 대한 평가 기준의 두 가지를 설정하였다. 정확도 측면에 대한 평가 기준은 기존의 추천 시스템 연구에서 보편적으로 사용한 방식을 채택하였다. 통계적인 정확도의 측정 기준으로 MAE를 사용하였다. 알고리즘의 성능에 대한 측정은 상당히 복잡한 문제이다. 실제 시스템의 성능은 알고리즘

의 특성과 데이터의 크기, 프로시저, H/W 및 S/W 사양 등의 변수에 영향을 받기 때문이다. 그러므로 우선 H/W나 S/W 사양에 영향을 받지 않는 시스템의 성능 평가 지표를 찾아볼 필요가 있다. 또한 알고리즘 자체의 성능 평가 지표를 도입할 수 있다면 그것을 통해 알고리즘의 성능의 우수성을 측정할 수 있을 것이다.

Agrawal[1]는 대규모 데이터베이스의 거래 데이터 속에서 아이템 간의 연관 규칙을 빠르게 발견하는 연구를 진행하였다. 이때 연관 규칙을 탐색하는 알고리즘의 성능을 평가하기 위한 방법으로 반응시간을 적용하였다. 즉, 분석 데이터의 크기가 변함에 따라 소요되는 탐색 시간을 상대화하여 나타냄으로써 알고리즘의 성능을 평가하였다. 본 연구에서도 이와 유사한 상황이므로 Agrawal의 연구에서 사용한 반응시간 협력적 필터링 개인화 기법의 성능을 측정하는 지표로 도입하였다. 본 연구에서도 예측이나 추천에 소요되는 반응시간을 측정함으로써 분석 데이터 사이즈(사용자의 수)의 변화에 따라 각각의 요소를 구현한 반응시간을 비교하였다. 여기서 선호도를 예측할 때 반응시간에 관한 문제가 존재한다. 하나의 텍스트일에 대한 선호도를 계산하기 위해서는 모든 사용자들과의 유사도를 계산하여야 하고 유사도를 바탕으로 다신 선호도를 계산해야 하므로 사용자가 전자 상거래 시스템에서 실 시간적으로 실행하기에는 많은 연산 시간을 요구한다. 이러한 반응 시간의 문제를 해결하기 위한 연구로써 텍스트일의 수를 줄여 유사도 계산의 시간을 줄이는 방법이 있고, 사용자의 수를 줄여 유사도 계산할 대상의 수를 줄이는 방법이 있다.

5. 실험 결과 및 해석

이웃 선택의 실험은 텍스트일에 대한 예측하기 위한 적정 이웃의 수를 10명으로 정하였다. 감성 데이터베이스에 있는 942명의 응답자들을 대상으로 100명씩 사용자 수를 증가시키면서 MAE와 반응시간을 측정하였다. 반응시간은 패션 디자인 추천 에이전트 시스템이 사용자에 적절한 텍스트일을 예측하는데 걸리는 시간(Millisecond)과 사용자가 추천된 텍스트일에 대해서 피드백을 하는 시간을 포함한다.

그림 5는 사용자 수에 따른 이웃 선택의 MAE와 반응시간의 결과를 나타낸다. 이웃 선택을 사용하였을 경우 성능 측면에서 상당한 향상이 있는 것으로 나타났다. 이웃 선택을 적용한 경우 사용자의 수를 증가시킬수록 점점 성능이 개선되었음을 알 수 있다. 이는 정확도 측면에서 기존의 연구들과 유사한 결과이다[21].

추천 텍스트일 집합의 실험은 942명의 사용자들이 평가한 1,240개의 텍스트일을 대상으로 진행하였다. 여기서 추천 대상은 예측 값이 높은 5개의 텍스트일을 추천

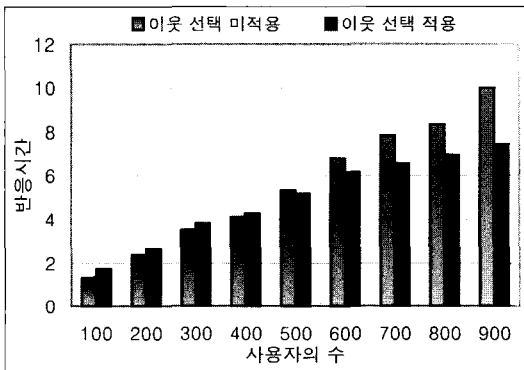
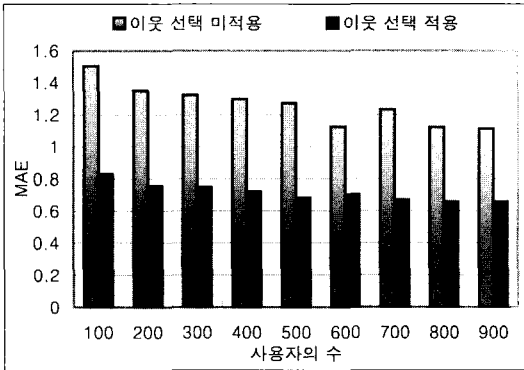


그림 5 이웃 선택에 대한 실험 결과

하였다. 추천 텍스타일 집합의 경우 추천을 제공하기 위해 예측해야 할 텍스타일의 수를 줄여주는 역할을 할 뿐, 예측 정확도에 영향을 미치는 것은 아니므로 반응시간만을 측정하였다. 그 결과 추천 텍스타일 집합을 적용했을 경우가 성능의 향상을 가져왔다. 추천 텍스타일 집합을 적용하지 않는 경우에는 텍스타일의 사이즈 증가에 비례하여 5개의 텍스타일을 추천하는데 소요되는 시간이 증가하였다. 그러나 추천 텍스타일 집합을 적용한 경우에는 반응시간이 증가하기는 했지만 그 크기는 미미하였다. 그림 6은 텍스타일 수에 따른 추천 텍스타일 집합에 대한 반응시간의 실험 결과를 나타낸다.

유사 군집의 실험은 942명의 사용들을 대상으로 하였다. 사용자의 수를 100명씩 증가시키면서 유사 군집의 중심 값과 그 군집에 속한 사용자의 실제 선호도의 차이로 MAE와 반응시간을 측정하였다. 실험에 포함된 사용자들의 희박성 문제를 해결하기 위해서 추천 에이전트 시스템이 선호도의 결측값에 대해서 예측한 값으로 대체하여 진행하였다. 유사 군집의 실험 결과는 성능과 정확도 사이의 상충 관계가 대비되어 나타났다. 성능 측면에서 사용자의 수에 관계없이 일정한 속도로 추천을 보여주는 반면 정확도의 경우에는 다소 손실을 보여주었다. 또한 사용자의 수와는 상관없이 일정한 정확도를 유

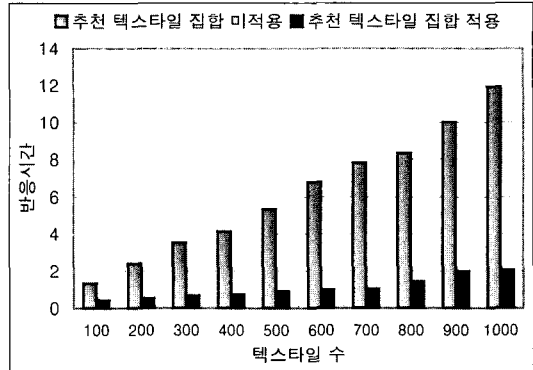


그림 6 추천 텍스타일 집합에 대한 실험 결과

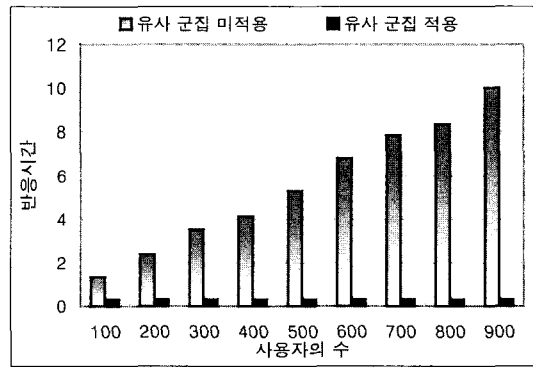
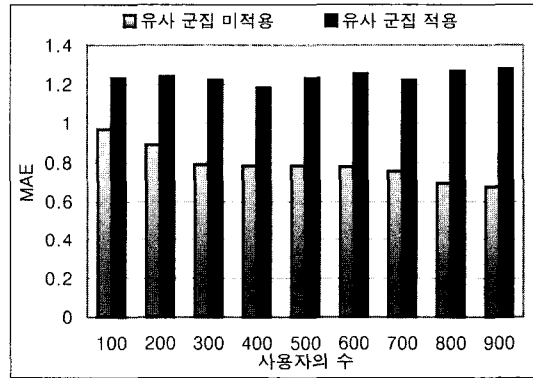


그림 7 유사 군집에 대한 실험 결과

지하였다. 이 결과를 통하여 유사 군집의 경우 데이터 사이즈의 증감보다는 유사 군집 자체의 문제가 정확도를 좌우하는 요소임을 알 수 있다. 그림 7은 사용자 수에 따른 유사 군집의 MAE와 반응시간의 실험 결과를 나타낸다.

6. 결론

협력적 필터링 개인화 기법을 이용한 소재 설계를 위한 텍스타일 감성 디자인 추천 에이전트 시스템의 성능

평가에 있어서 정확도와 성능 문제가 있다. 즉 데이터의 규모가 커질수록 예측이 정확도는 향상되는 반면, 성능은 급격히 저하된다. 전자상거래 사이트에서 실제로 유용한 추천 시스템을 구축하기 위해서는 성능과 정확도 문제를 동시에 고려하는 것이 중요하다. 그러나 기존의 연구에서는 성능 문제를 배제하고 예측의 정확도에 초점을 두어 연구되어 왔다. 본 연구는 협력적 필터링 개인화 기법을 이용한 추천 시스템의 성능과 정확도를 향상시킬 수 있는 방안에 관한 연구를 하였다. 즉, 추천 시스템의 성능에 초점을 두고 있는 이웃 선택, 추천 텍스타일 집합, 군집 분석을 패션 디자인 추천 에이전트 시스템에 추가하였다. 구체적으로, 협력적 필터링 개인화 기법을 이용한 패션 디자인 추천 에이전트 시스템을 실제로 구현하고, 정확도 측면에서 MAE, 성능 측면에서 반응시간을 평가 지표로 사용하였다. 결과적으로, 이웃 선택은 정확도와 성능을 모두 향상시킨 것으로 나타났다. 추천 텍스타일 집합의 경우 예측의 정확도에 영향을 주지 않으면서 텍스타일 추천을 산출하고 반응시간을 단축시켜 시스템의 효율성을 향상시킬 수 있었다. 유사 군집의 경우 사용자의 수에 관계없이 성능을 증가시킨 반면 정확도는 다소 낮아진다. 이와 같이 개인화 텍스타일 감성 디자인 추천시스템의 추천의 정확도와 성능 문제를 동시에 고려하였다는 점에서 중요하다고 할 수 있다.

참고 문헌

- [1] R. Agrawal and T. Imielinski and A. Swami, "Mining association rules between sets of items in large databases," In Proc. of the 1993 ACM SIGMOD Conference, Washington DC, USA, 1993.
- [2] P. Resnick, et. al., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," In Proc. of ACM CSCW'94 Conference on Computer Supported Cooperative Work, pp. 175-186, 1994.
- [3] J. Herlocker, J. Konstan, A. Borchers and J. Riedl, "An Algorithm Framework for Performing Collaborative Filtering," In Proc. of ACM SIGIR'99, 1999.
- [4] N. Good, B. Schafer, J. Konstan, A. Borchers, B. Sarwar, J. Riedl, "Combining Collaborative filtering with Personal Agents for Better Recommendation," AAAI/IAAI, 1999.
- [5] J. S. Breese and D. Heckerman and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," In Proc. of the 14th Conference on Uncertainty in Artificial Intelligence, 1998.
- [6] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Analysis of Recommender Algorithms for E-Commerce," ACM E-Commerce Workshop, 2000.
- [7] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Application of Dimensionality Reduction in Recommender System," ACM WebKDD 2000 Web Mining for E-Commerce Workshop, 2000.
- [8] 윤영미, 정경용, "협력적 필터링 기술에서 평균 정보량인 엔트로피를 이용한 효율적인 예측 방법", 한국정보기술학회논문지, 제1권, 제1호, pp.73-83, 2003.5.
- [9] K. Y. Jung, Y. J. Na, J. H. Lee, "FDRAS: Fashion Design Recommender Agent System using the Extraction of Representative Sensibility and the Two-Way Filtering on Textile," Lecture Notes in Computer Science 2736, Springer-Verlag, pp.631-640, 2003.
- [10] K. Y. Jung, J. J. Jung, J. H. Lee, "Discovery of User Preference in Personalized Design Recommender System through Combining Collaborative Filtering and Content Based Filtering," Lecture Notes in Artificial Intelligence 2843, Springer-Verlag, pp.320-327, 2003.
- [11] 정경용, 나영주, 이정현, "Textile 기반의 협력적 필터링 기술을 이용한 섬유 패션 디자인 추천 시스템 개발", 정보과학회논문지: 컴퓨팅의 실제, 제9권, 제5호, pp.541-550, 2003.10.
- [12] K. Y. Jung, K. W. Rim, J. H. Lee, "Automatic Preference Mining through Learning User Profile with Extracted Information," Lecture Notes in Computer Science 3138, Springer-Verlag, pp.815-823, 2004.
- [13] K. Y. Jung, Y. J. Na, D. H. Park, J. H. Lee, "Discovery Knowledge of User Preferences: Ontologies in Fashion Design Recommender Agent System," Lecture Notes in Computer Science 3044, Springer-Verlag, pp.863-872, 2004.
- [14] K. Y. Jung, Y. J. Na, J. H. Lee, "Creating User-Adapted Design Recommender System through Collaborative Filtering and Content Based Filtering," Lecture Notes in Artificial Intelligence 2902, Springer-Verlag, pp.204-208, 2003.
- [15] K. Y. Jung, J. H. Lee, "Prediction of User Preference in Recommendation System using Association User Clustering and Bayesian Estimated Value," Lecture Notes in Artificial Intelligence 2557, Springer-Verlag, pp.284-296, 2002.
- [16] 권규식, 이정우, "감성공학적 제품개발을 위한 감성과 디자인 요소와의 관계 모형화", 한국감성과학회 연차 학술대회, pp.11-15, 1997.
- [17] 이순요, 양선모, 감성 공학, 청문각, 1996.
- [18] 황상민, 정지홍, 정선화, 김지연, "소비자 선호와 라이프 스타일에 기초한 컨셉트 제품의 감성 이미지 분석", 한국정보과학회 HCI·CG·DESIGN 학술대회는문집(1), pp.559-607, 2003.2.
- [19] 이주현, "섬유 디자인에 대한 감성 공학적 연구 사례", 섬유기술과 산업, 제2권, 제4호, 1998.
- [20] 홍태화, 협력적 필터링 알고리즘을 이용한 추천 시스템의 퍼포먼스 향상 방안에 관한 연구, 서울대학교 대학원 석사학위논문, 2000.

- [21] 고수정, 전자상거래에서 협력적 여과와 내용 기반 여과를 병합한 사용자 선호도 마이닝. 인하대학교 대학원 박사학위논문, 2002.
- [22] 정경용, 협력적 여과 시스템에서 연관 사용자 군집과 베이지안 추정치를 이용한 예측 방법, 인하대학교 대학원 석사학위논문, 2002.



정 경 용

2000년 인하대학교 전자계산공학과(공학사), 2002년 인하대학교 전자계산공학과(공학석사), 2002년~현재 인하대학교 전자계산공학과 박사과정. 2001년~현재 에이플러스전자(주) 선임연구원. 2003년~현재 가천길대학 컴퓨터소프트웨어과 겸임교수. 관심분야는 웹 마이닝, 기계학습, 정보검색, CRM, 협력적 필터링, 자연어처리, 전자상거래



김 중 훈

2001년 인천대학교 물리학과(이학사) 2003년 인하대학교 전자계산공학과(공학석사), 2004년~현재 인하대학교 전자계산공학과 박사과정. 2005년~현재 가천길대학 컴퓨터소프트웨어과 겸임교수. 관심분야는 정보검색, 자연어처리, 음성신호처리, 임베디드시스템



나 영 주

1988년 서울대학교 의류학과(이학사) 1990년 서울대학교 대학원(이학석사) 1994년 University of Maryland Textiles & Consumer economics PhD 1995년 서울대학교 의류학과 시간강사 1995년~현재 인하대학교 의류학과 교수 관심분야는 감성과학, 텍스트일정보 및 기획, 의복환경



이 정 현

1977년 인하대학교 전자공학과 졸업 1980년 인하대학교 대학원 전자공학과(공학석사), 1988년 인하대학교 대학원 전자공학과(공학박사), 1979년~1981년 한국전자기술연구소 시스템 연구원. 1984년~1989년 경기대학교 전자계산학과 교수. 1989년~현재 인하대학교 컴퓨터공학부 교수. 관심분야는 자연어처리, HCI, 정보검색, 음성인식, 음성합성, 컴퓨터구조