

협업 필터링과 개인 성향을 이용한 여성 의류 온라인 쇼핑몰 추천 시스템

신해란*, 김승언*, 박두순*

*순천향대학교 컴퓨터소프트웨어공학과
e-mail:98gofks@naver.com , tmddjs1127@naver.com

A recommendation system for women's clothing online shopping mall using collaborative filtering and personal propensity

Hae-Ran Shin, Seong-Eon Kim, Doo-Soon Park**

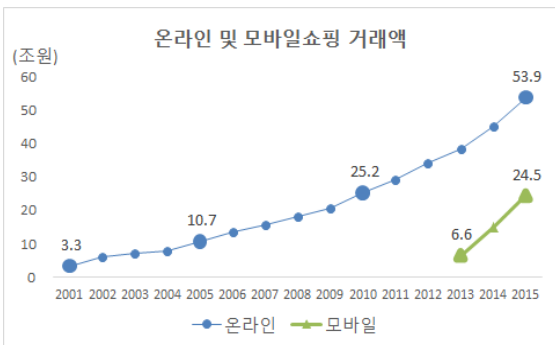
*Dept. of Computer Software Engineering, SoonChunHyang University

요 약

최근 스마트폰의 보급률이 높아지면서 인터넷 쇼핑물의 접근성이 용이해지고 있고 그로 인해 사용자들의 인터넷 쇼핑의 이용이 보편적이게 되었다. 그 중 여성 의류 분야는 많은 비중을 차지하고 있으며 현재도 꾸준히 성장하고 있는 추세이다. 많은 여성 소비자들은 개인의 취향에 맞는 의류들을 추천받기를 원한다. 본 논문에서는 협업 필터링에서 발생하는 cold start 문제를 이름, 나이, 선호 스타일, 자주 사용하는 쇼핑물 등 개인 성향을 이용하여 해결하는 협업 필터링과 개인 성향을 이용한 여성 의류 쇼핑몰 추천 시스템을 제안한다.

1. 서론

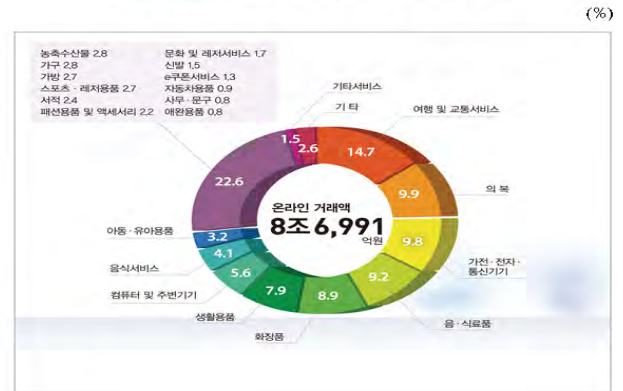
최근 과학기술의 발전으로 인해 스마트폰의 보급률이 높아지면서 많은 스마트폰 사용자들이 시간과 장소에 구애받지 않으며 인터넷을 사용할 수 있게 되었다. 이에 따라 인터넷 쇼핑물의 접근이 용이해지면서 인터넷 쇼핑물의 거래량이 대폭 상승하였고 그에 따라 거래액도 증가하였다. (그림 1)은 최근 20년간의 온라인 및 모바일 쇼핑 거래액이다.



(그림 1) 온라인 및 모바일쇼핑 시장 상승 추이[1]

(그림 2)는 온라인 쇼핑 거래액 상품군 별 구성비를 나타낸 것이다.

< 상품군별 온라인쇼핑 거래액 구성비 >



(그림2) 2018년 1월 국내 온라인 쇼핑 동향[2]

쇼핑몰 솔루션 사이트인 메이크샵의 통계에 따르면 여성 의류 온라인쇼핑몰의 개인 쇼핑물 점유율은 25.8%로 6.9%미만인 남성 의류 쇼핑몰 보다 현저하게 많은 상태이다. 이처럼 여성 의류 쇼핑몰의 수요가 많은 상태에서 소비자들은 자신의 패션 스타일과 취향에 맞는 온라인 쇼핑물을 선별하여 이용할 필요가 있다.

이러한 요구 사항들을 충족시켜주기 위하여 협업필터링을 이용한다. 협업필터링은 데이터가 없는 서비스 초기에는 추천이 부정확하다는 Cold start 문제가 발생한다. 이를 해결하기 위해서 본 논문에서는 협업 필터링 방법과

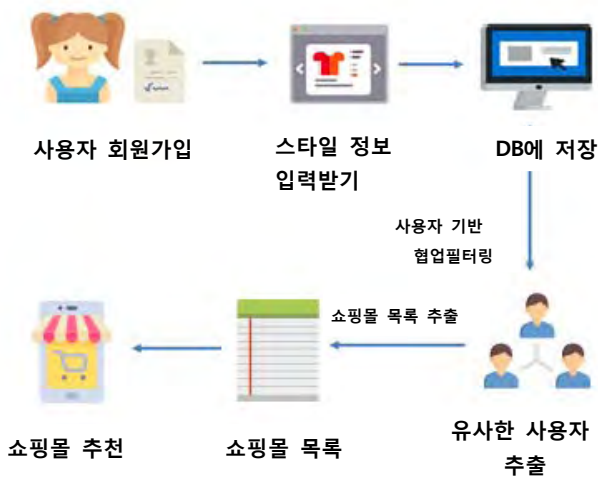
※ "본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업의 연구결과로 수행되었음" (IITP-2018-2014-1-00720)

이름, 나이, 선호 스타일, 자주 사용하는 쇼핑몰 등의 사용자 개인 성향을 이용하여 개인 맞춤형 쇼핑몰을 추천하는 시스템을 제안한다.

2. 쇼핑몰 추천 시스템의 구성

협업 필터링이란, 사용자들의 선호도와 관심 표현을 바탕으로 선호도, 관심도가 비슷한 사용자들을 식별해 내는 방법으로 과거에 이용한 콘텐츠가 비슷하다면 사용자 간에 유사한 성향을 가지고 있다고 판단하고 그 근거를 토대로 추천하는 방식이다[3]. 추천 시스템은 사용자의 과거 선호도에 기반 하여 예측한 후 새로운 아이템에 대한 평가치를 예측하여 제공하는 시스템이다[4]. 새로운 사용자들의 정보가 없을 때 Cold start 문제가 발생하는데 이를 해결하기 위하여 신규 사용자는 이름, 나이, 평소 선호하는 스타일, 자주 사용하는 쇼핑몰 사이트 명 등 개인 성향을 입력받는다. 그리고 유사성의 정확도를 높이기 위하여 상의, 아우터, 하의, 신발의 세부항목들을 질문한다.

본 논문에서 제시하면 추천 시스템의 시나리오는 (그림 3)과 같다.



(그림 3) 추천 시스템 시나리오

(그림 3) 과 같은 쇼핑몰 추천 시스템의 시나리오에서는 협업필터링을 통해 유사한 스타일을 선호하는 사용자들이 이용한 쇼핑몰들을 목록을 추출하게 된다. 신규 사용자는 이름, 나이 등의 기본적인 정보를 수집하고 평소 선호하는 스타일과 자주 사용하는 쇼핑몰 사이트 명을 입력받는다. 그리고 유사성의 정확도를 높이기 위하여 상의, 아우터, 하의, 신발의 세부항목들을 질문한다. 새로운 사용자가 들어오면 협업 필터링과 개인 성향을 이용하여 유사도를 계산한다. 유사도는 피어슨 상관계수를 이용하였다. 피어슨 상관계수는 -1부터 1의 값을 가지며, 관계가 크면 1에 가까워지고 서로의 유사함이 적으면 -1에 가까워진다. 피어슨 상관계수를 이용한 수식은 아래 (그림 4)와 같다.

$$r = r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

(그림 4) 피어슨 상관계수 공식

피어슨 상관계수 공식에서 r_{xy} 는 사용자 x 와 사용자 y 사이의 유사도이고 x_i 와 y_i 는 각각의 사용자가 아이템 i 에 대하여 평가한 값이며 \bar{x} 와 \bar{y} 는 각각의 사용자가 과거에 제공한 모든 평가들의 평균이다[5].

유사도 측정이 끝나면 유사도가 높은 순서로 Top-2리스트를 추출한다. Top-2리스트에 있는 기존 고객들이 추천하는 쇼핑몰을 새로운 사용자에게 추천한다.

3. 쇼핑몰 추천 시스템의 구현

회원가입 양식은 (그림 5)과 같다. 회원가입 양식을 작성한 후 스타일 정보를 수집하는 양식은 (그림 6)과 같다.

(그림 5) 회원가입 양식

Style 정보 수집

(그림 6) Style 정보 수집 양식

(그림 6)의 Style 정보 수집 양식에서 실제 style 선택지는 총 15가지가 있고 2가지의 항목을 복수 선택할 수 있으며

자주 찾는 쇼핑물 사이트는 반드시 2개를 입력해야 한다. 또한 상의, 아우터, 하의, 신발의 정보도 입력받아 유사성의 정확도를 높인다. 입력한 정보들은 Data Base(이하 DB)에 저장되며 DB의 일부는 (표 1)과 같다.

(표 1) 사용자 DB의 일부분1

| User | | | | | | | | |
|---------|-----|----|------------|-------------|------|-----|----|-----|
| ID | 이름 | 나이 | 선호 스타일 | 쇼핑물 | 상의 | 아우터 | 하의 | 신발 |
| User1 | 김개똥 | 27 | 리블리,유니크 | 언니가간다,육육걸즈 | 후드티 | 야상 | 바지 | 구두 |
| User2 | 최나나 | 18 | 심플메이직,아메카지 | 육육걸즈,햇핑 | 니트 | 잠퍼 | 바지 | 운동화 |
| User3 | 박미미 | 31 | 섹시글램,로맨틱 | 히프나틱,뽀뽀데이즈 | 블라우스 | 자켓 | 치마 | 구두 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| User496 | 오다다 | 19 | 유니크,로맨틱 | 언니가간다,데쥬 | 후드티 | 코트 | 치마 | 구두 |
| User497 | 박라라 | 23 | 리블리,유니크 | 쥬,데쥬 | 셔츠 | 야상 | 바지 | 단화 |
| User498 | 최아하 | 21 | 유니크,캠피스 | 스타일난다,프롬비기닝 | 맨투맨 | 가디건 | 바지 | 운동화 |
| User499 | 홍미나 | 20 | 섹시글램,모던시크 | 프롬데이즈,스타일난다 | 블라우스 | 패딩 | 치마 | 구두 |
| User500 | 김다라 | 36 | 레미닌,오피스 | 임블리,히프나틱 | 셔츠 | 자켓 | 바지 | 구두 |

본 논문에서는 사용자가 회원가입 할 때 입력한 각각 선호style, 선호쇼핑물과 상의, 아우터, 하의, 신발의 총 6 가지 항목을 이용해 신규 사용자의 추천 DB를 구축한다. 이러한 데이터를 토대로 사용자와 다른 사용자간의 유사도를 피어슨 상관계수 (Pearson product-moment correlation coefficient)를 통해 구한다. 유사도를 결정하는 각 항목에 대하여 피어슨 상관계수로 계산을 하고 각 항목의 값에 대한 평균을 두 사용자 간에 비교했을 때 가장 높은 유사도를 가지고 있는 사용자의 쇼핑물이 추천된다. 이러한 피어슨 상관계수를 이용하여 User1에 대한 사용자의 유사도 측정을 실시하였다. User1의 이름은 김개똥, 나이는 27살이다. User1이 선호하는 스타일과 자주 사용하는 쇼핑물, 상의, 아우터, 하의, 신발의 정보는 (표 1)과 같다.

(표 2)는 User1에 대해 사용자의 유사도 측정을 한 결과에 대해 일부를 나타낸 표이다.

(표 2) User1을 기준으로 유사도를 측정한 표

| ID | Compare User | Similarity |
|---------------|--------------|------------|
| My ID : User1 | ID:User2 | 0.469 |
| | ID:User3 | 0.028 |
| | ⋮ | ⋮ |
| | ID:User496 | 0.684 |
| | ID:User497 | 0.981 |
| | ID:User498 | 0.287 |
| | ID:User499 | 0.631 |
| ID:User500 | 0.247 | |

(표 2)를 정렬하여 유사도가 가장 높은 추천받을 사용자와 가장 유사한 사용자 2명을 추출하면 (표 3)와 같다.

(표 3) User1과 가장 유사한 2명의 쇼핑물 추천 리스트

| 쇼핑물 추천 | ID:User496 | 언니가간다,데쥬 |
|--------|------------|----------|
| | ID:User497 | 쥬,데쥬 |

이렇게 해서 추천되는 쇼핑물은 언니가 간다 1번, 데쥬 2번, 쥬 1번이다. 이를 토대로 데쥬, 언니가 간다, 쥬 순서로 추천해야 하지만 User1이 언니가 간다, 육육걸즈를 추천하고 있으므로 언니가 간다를 제외한 데쥬, 쥬가 추천된다. 이렇게 추천된 쇼핑물은 사용자에게 쇼핑물의 로고, 이름과 사이트 링크가 게시된 추천 결과 화면을 보여준다. (그림 7)은 User1의 추천 결과이다.



(그림 7) 추천 결과 화면1

같은 방식으로 User78에 대해서도 추천을 한다. (표 4)는 사용자 DB의 일부분이고 (표 5)는 User78을 기준으로 유사도를 측정한 표이다.

(표 4) 사용자 DB의 일부분2

| User | | | | | | | | |
|---------|-----|----|-----------|--------------|------|-----|----|-----|
| ID | 이름 | 나이 | 선호 스타일 | 쇼핑물 | 상의 | 아우터 | 하의 | 신발 |
| User31 | 임규규 | 22 | 섹시글램,레미닌 | 언니가간다,육육걸즈 | 니트 | 코트 | 치마 | 구두 |
| User32 | 이수수 | 23 | 스물륙,아메카지룩 | 파라노말샵,언니가 간다 | 니트 | 패딩 | 바지 | 운동화 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| User78 | 김지지 | 20 | 스물륙,캠피스룩 | 멜로우루피스,언니룩 | 블라우스 | 패딩 | 바지 | 운동화 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| User263 | 윤정정 | 16 | 스물륙,캠피스룩 | 언니가 간다,소퍼랜드 | 블라우스 | 패딩 | 바지 | 운동화 |
| User264 | 신희희 | 33 | 리블리,유니크 | 임블리,왕투 | 셔츠 | 자켓 | 치마 | 구두 |

(표 5) User78을 기준으로 유사도를 측정한 표

| ID | Compare User | Similarity |
|----------------|--------------|------------|
| My ID : User78 | ID:User31 | 0.231 |
| | ID:User32 | 0.896 |
| | ⋮ | ⋮ |
| | ID:User100 | 0.102 |
| | ID:User101 | 0.109 |
| | ⋮ | ⋮ |
| | ID:User263 | 0.998 |
| | ID:User264 | 0.377 |

(표 5)를 정렬하여 유사도가 가장 높은 추천받을 사용자와 가장 유사한 사용자 2명을 추출하면 (표 6)와 같다.

(표 6) User78과 가장 유사한 2명의 쇼핑물 추천 리스트

| | | |
|--------|------------|---------------|
| 쇼핑물 추천 | ID:User32 | 파라노말샵, 언니가 간다 |
| | ID:User263 | 언니가 간다, 쇼퍼랜드 |

이렇게 해서 추천되는 쇼핑물은 파라노말샵 1번, 언니가 간다 2번, 쇼퍼랜드 1번이다. 이를 통해 구현된 결과 화면은 (그림 8)과 같다.



(그림 8) 추천 결과 화면2

4. 결론

본 논문에서는 쏟아지는 쇼핑물 사이트들 사이에서 사용자에게 가장 적합한 쇼핑물 사이트를 추천하기 위해 협업 필터링을 기반으로 한 프로그램을 제작했다. 사용자의 데이터를 토대로 피어슨 상관계수를 이용해 사용자의 유사도를 측정하였고 이 결과를 토대로 추천 결과화면을 구현하였다. 향후 연구과제는 사용자의 데이터베이스를 구성하는 요소들의 개수를 추가하여 보다 정확한 유사성을 추출해 내는 것이다.

참고문헌

- [1] “통계로 본 온라인쇼핑 20년”, 통계청, 2016
- [2] “2018년 1월 온라인쇼핑동향”, 통계청, 2018
- [3] 김영아, 박두순, “협업 필터링 기반 드라마 추천 시스템”, 한국정보처리학회 춘계학술대회 발표 논문집, 제주한라대학교, pp. 1137-1138, 2013.11
- [4] 김경록, 변재희, 문남미, “장르유사도와 선호장르를 이용한 협업필터링 설계”
- [5] 조진관, “피어슨 상관계수 측정을 이용한 추천 알고리즘의 성능 비교분석”