

# 협업필터링과 개인 식단 선호도를 이용한 다이어트 식단 추천 시스템

\*장태훈, \*박두순  
\*순천향대학교 컴퓨터소프트웨어공학과  
e-mail : wkdxogns23@naver.com

## Diet menu recommendation system using collaborative filtering and personal menu preference

Tae-Hoon Jang, Doo-Soon Park\*  
Dept. of Computer Software Engineering, Soonchunghyang University

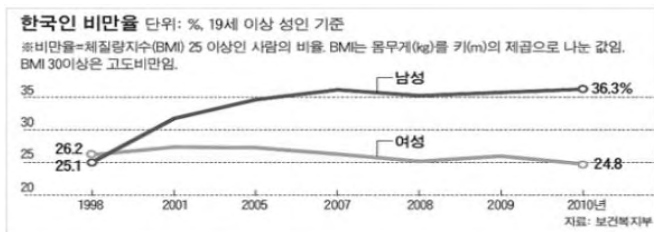
### 요 약

최근 서구화된 식습관으로 인해 국내 비만 인구가 급격하게 증가했고 이에 따라서 다이어트에 관심을 갖는 사람들이 많아졌다. 하지만 현재 출시되어 있는 애플리케이션은 단지 식단을 선택하도록 목록을 보여주는 것에 그친다. 최근 추천 시스템은 데이터 서비스에 필수적인 요소가 되었고 사용자가 직접 선택에 대한 고민 할 것 없이 개인에게 맞춤형 식단을 제공 할 필요가 있다. 따라서 본 논문에서는 FatSeceret Open-API를 사용하여 추출한 식품 데이터베이스를 통해 사용자의 식단 선호도를 분석하였다. 그래서 식단 선호도와 협업 필터링 기법을 R을 이용하여 개인 맞춤형 다이어트 식단을 추천해주는 시스템을 개발한다.

키워드 : 협업 필터링, 다이어트 식단

### 1. 서론

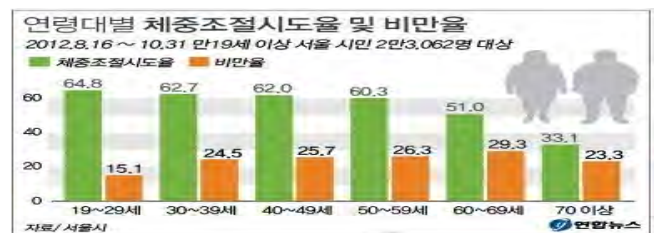
비만이란 체내에 지방조직이 과다한 상태를 말하며 최근 서구화된 음식들이 국내에 많이 유입됨에 따라 비만율이 급격하게 증가하는 추세를 보인다. 보건복지부의 조사에 따르면 2010년 한국 남성의 36.3%, 여성은 24.8이 비만이라는 통계를 볼 수 있을 정도로 한국인의 비만율은 사상최고를 기록하고 있다[1]. 1998년 ~ 2010년까지의 한국인 비만율은 (그림 1)과 같다.



(그림 1) 1998년 ~ 2010년 한국인 비만율[1]

비만은 많은 합병증을 유발하며 대표적으로 고지혈증, 당뇨병, 관절염, 심혈관계 질환을 들 수 있고 비만 인구의

증가와 더불어 이러한 합병증을 앓는 인구 또한 증가하고 있다. 그로인해 최근 다이어트에 대한 관심은 매우 증가했으며 다이어트를 시도하는 인구 또한 늘어나고 있다. 연합뉴스에서 조사한 연령대별 체중조절 시도율 및 비만율은 (그림 2)와 같다.



(그림 2) 연령대별 체중조절시도율 및 비만율 [2]

비만을 예방 하는데 있어 식이요법으로 하루 총 섭취 칼로리에서 500 Kcal을 줄이면 체중이 1주일에 약 0.5kg을 감소할 수 있을 정도로 식이요법은 중요하다[3]. 하지만 자신에게 적합한 다이어트 식단을 찾는 것은 쉬운 일이 아니다. 현재 ‘다이어트 코치’, ‘먹고 마시고’와 같은 다이어트에 도움이 되는 애플리케이션이 출시되어 있지만 이러한 애플리케이션은 단지 자신이 먹은 식단을 기록 할 수 있고 몇 가지 다이어트 식단들을 보여주는 것이 전부이기 때문에 사용자들은 식단을 추천 받는데 어려움을 겪고 있다.

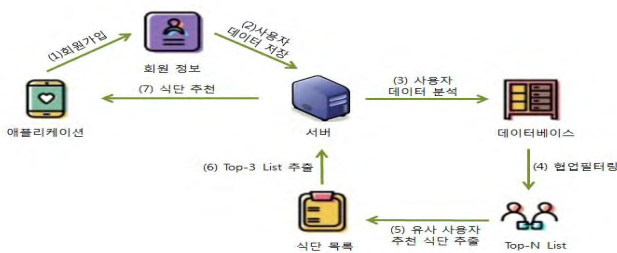
※ “본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음” (IITP-2018-2014-1-00720)

최근 추천은 데이터 서비스에서 필수적인 요소가 되어있다. 이는 너무나도 많은 정보들이 생성됨에 따라, 서비스를 이용하는 사람들은 다양한 데이터에 둘러싸여 사용자가 필요한 데이터를 보고 직접 찾거나 선택하는 것보다 서비스에서 직접 어떠한 선택에 대한 고민할 필요 없이 관심이 있거나 유용한 정보를 주는 방식을 선호한다[4]. 따라서 본 논문에서는 편리한 다이어트 식단 선택을 돕기 위해 Open-API를 활용하여 음식 데이터를 수집하며 개인화 요소를 바탕으로 협업필터링을 이용해 사용자의 건강과 기호에 적합한 식단을 추천하는 시스템을 제안한다.

## 2. 시스템의 구성

본 논문에서는 추천을 제공하기 위해 협업 필터링 기술을 사용한다. 협업 필터링은 기존 사용자의 선택 이력을 바탕으로 하여 미래의 선택 이력 또한 과거와 비슷한 성향을 띄울 것이라는 가정 하에, 사용자와 유사한 취향을 가진 사용자들을 유사 집단으로 분류하고 사용자가 아직 선택하지 않은 유사 집단의 추천 상품을 사용자에게 추천해주는 방법을 말한다[5].

본 논문에서 구현한 추천 시스템 시나리오는 (그림 3)과 같다.



(그림 3) 추천 시스템 시나리오

(그림 3)의 추천 시스템 시나리오의 설명은 다음과 같다. (1) 먼저 사용자는 추천 시스템을 사용하기에 앞서 회원가입을 통해 자신의 나이, 체중, 키, 성별, 추천 식단을 입력한다. 이 때 추천 식단은 사용자의 선호 식단 정보를 습득하는데 사용된다. (2) 회원 가입 정보는 회원 데이터베이스에 저장되며 (3) 이렇게 가입된 사용자 정보와 사용자 추천 식단은 기존에 가입되어 있는 다른 사용자들 사이의 유사도 측정을 통해 그룹화 되어있는 그룹들과 협업필터링을 통해 유사도를 측정하는데 사용된다. (4) 사용자와 사용자 그룹들 사이의 관계는 R에서 지원하는 'recommendarlab'를 이용하여 Jaccard Similarity 공식을 통해 유사도를 측정하고 협업 필터링을 이용해 사용자와 유사한 유사 그룹 Top-N List를 구성한다. (5) Top-N List는 사용자와 유사도가 높은 사용자 집단을 말하며 Top-N List의 사용자들이 추천한 식단을 사용자에게 추천한다. (6) 이렇게 추출된 Top-N List가 추천한 음식은 Open-API에서 수집한 데이터베이스를 통해서 성분과 브랜드 등의 (표 2)의 기준으로 분석하고 다이어트 식단이 맞는지를 판독하며 판독된 식단 중 유사 그룹에서 가장

많이 추천한 음식과 사용자 본인이 추천하지 않은 식단을 Top-3 List로 구성한다. (7) 이러한 방법으로 Top-3 List가 추천한 식단을 사용자에게 추천하는 시나리오를 갖는다.

본 논문에서 분류한 개인화 요소는 비만도 진단법을 기준으로 소분류 했고 사용자 분석 데이터로 사용 될 개인화 요소들은 키, 나이, 체중, 성별로 (표 1)과 같다

(표 1) 개인화 요인 사용 요소 [6]

번호	키(cm)	나이	체중(kg)	성별
1	120 ~ 130	10대	50 ~ 60	남성
2	130 ~ 140	20대	60 ~ 70	여성
3	140 ~ 150	30대	70 ~ 80	
4	150 ~ 160	40대	80 ~ 90	
5	160 ~ 170	50대	90 ~ 100	
6	170 ~ 180	60대	100 ~ 110	
7	180 ~ 190	70대	110 ~ 120	
8	190 ~ 200	80대	120이상	

(표 1)의 개인화 요소는 회원가입을 할 때 필수적으로 입력 받으며 사용자가 추천하고 싶은 식단 5개를 입력한다. 이 개인화 요소는 협업필터링과 유사도 측정에 사용된다.

## 3. 시스템의 구현

본 논문에서는 키, 나이, 체중, 성별 등 비만도 측정 요소와 키, 나이, 체중, 성별, 직업, 활동량 등 통계청 대분류를 참고하여 구성한 개인화 요소를 9000명의 사용자 데이터를 토대로 바꿔가며 추천해본 결과 비만도 측정 요소를 사용했을 때와 통계청 대분류의 요소들을 포함하여 추천했을 때의 결과가 매우 유사하게 나타났다. 때문에 비만도 측정 요소인 키, 나이, 체중, 성별을 제외한 통계청 대분류의 개인화 요소들은 제외하도록 했다. (그림 4)는 비만도 측정에 필요한 요소들을 통해 측정된 유사도 측정 결과다. (그림 5)는 통계청 대분류를 포함하여 측정된 유사도 측정 결과 중 가장 유사도가 높게 측정된 결과로 (그림 4)의 요소에서 직업, 활동량을 포함한 결과의 일부이다.

ID	Filter similarity
1	0.723
2	0.678
3	0.665
4	0.542
5	0.524
6	0.533
7	0.521
8	0.478
9	0.441
10	0.422

(그림 4) 키, 나이, 체중, 성별을 이용한 유사도 측정 결과

ID	Filter similarity
1	0.721
2	0.666
3	0.645
4	0.532
5	0.524
6	0.511
7	0.423
8	0.412
9	0.409
10	0.399

(그림 5) 키, 나이, 체중, 성별, 직업, 활동량을 이용한 유사도 측정 결과

본 논문에서는 협업필터링을 통해 추출된 사용자와 유사한 집단인 Top-N List가 추천한 식단을 사용자에게 추천하기에 앞서 추천 식단을 분석하기 위해 Open-API로 식단 데이터베이스를 추출했다. Open-API는 정보를 얻어 새로운 서비스를 창출하기가 용이한 것이 특징이다. 이러한 Open-API 중 FatSeceret에서는 음식 칼로리와 1인분 분량의 무게, 성분정보, 브랜드 등을 데이터화하여 가지고 있으며 5000여개의 음식 정보를 API서비스로 얻을 수 있도록 무료 제공하고 있다. 따라서 본 논문에서는 필요한 정보를 FatSeceret에서 제공한 Open-API를 활용하여 수집했다[5].

(그림 6)은 Open-API를 통해 수집한 식품 데이터베이스의 일부이며 (표 2)는 FatSeceret에서 수집 가능한 데이터 목록이다.

역청	스위트플러도넛초 볼떡키리 (127kcal)	역청	크림치즈 크래들 (444kcal)	역청	자몽 3개 / 계란 1개 / 토마토 1 개 / 올리브기 름 1스푼 (275kcal)	역청	자몽 3개 / 계란 2개 / 토마토 1 개 / 올리브기 름 1스푼 (275kcal)
원성	치킨스킬릭 야채 볶음밥 (124kcal)	원성	보통안심 짜장밥 (213kcal)	원성	계란 3개 / 토마 토 1개 / 올리브 기름 1스푼 (184kcal)	원성	계란 3개 / 토마 토 1개 / 올리브 기름 1스푼 (184kcal)
치네	애쉬리얼 블루프 리니니(머스트드 드레싱)(213kcal)	치네	쉬리프호박로콜 리 블루프(1개분 소량)당유유 침(219kcal)	치네	계란 3개 / 야채 샐러드(200kcal)	치네	쇠고기 스테이크 / 토마토 / 샐러 드 / 올리브기름 (220kcal)
역청	크림치즈 샌드위치 (279kcal)	역청	애니언황치 샌드 위치(200kcal)	역청	자몽 3개 / 계란 1개 / 올리브기 름 1스푼 (275kcal)	역청	자몽 3개 / 계란 2개 / 토마토 1 개 / 올리브기 름 1스푼 (275kcal)
원성	향저야채 버빙밥 (122kcal)	원성	계몽말가슴살 매 추리탕 볶음 (150kcal)	원성	자몽 3개 / 계란 2개 / 토마토 1 개 / 올리브기 름 1스푼 (275kcal)	원성	자몽 3개 / 계란 2개 / 토마토 1 개 / 올리브기 름 1스푼 (275kcal)
치네	양파버섯로망 샐러드(1개분 소 량)당유유침(219kcal)	치네	양두부 샐러드 (당유유침) (150kcal)	치네	자몽 3개 / 계란 1개 / 토마토 1 개 / 올리브기 름 1스푼 (275kcal)	치네	자몽 3개 / 계란 2개 / 토마토 1 개 / 올리브기 름 1스푼 (275kcal)

(그림 6) 식단 데이터베이스의 일부

(표 2) FatSeceret에서 수집 가능한 데이터 목록[4]

데이터 타입	설명
food_type	브랜드제품인지 일반제품인지 여부
brand_type	제조, 레스토랑, 슈퍼마켓과 같은 생산 종류
calories	칼로리
carbohydrate	탄수화물
protei	단백질
fat	지방
trans_fat	트랜스지방
cholestero	콜레스테롤
sodium	나트륨
potassium	칼륨
fiber	섬유질
vitamin_a	2000칼로리 기준 일일권장비타민 a비율
vitamin_c	2000칼로리 기준 일일권장비타민 c비율
calcium	칼슘
iron	철
weight_type	측정 단위별 무게

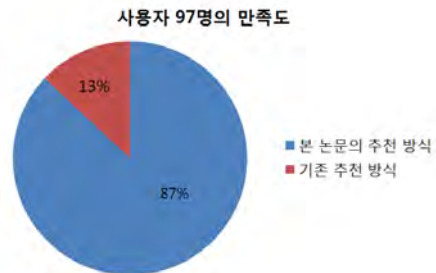
기존의 사용자들은 서로의 유사도 측정을 통해서 N개의 유사 그룹을 형성하고 있다. 그래서 새로운 사용자가 회원가입을 하면 이 그룹들과 사용자 사이에 유사도 측정을 통해서 유사 그룹인 Top-N List를 추출하는데 이때 Top-N List의 사용자들이 추천한 식단 중 가장 많이 추천 되었으며 추천을 받는 사용자가 추천하지 않은 식단을 Top-3 List로 추출한다. 하지만 이 때 추천 받는 식단이

다이어트에 적합하지 않은 식단인 경우들이 포함되어 있었다. 그래서 본 논문에서는 Top-N List가 추천한 식단과 Open-API로 제공 받은 식품 데이터베이스를 (표 2) 기준으로 비교 분석하며 이를 통해서 사용자에게 적합한 다이어트 식품인지를 찾는다. (표 3)은 User1의 Top-N list가 추천한 식단을 분석하여 다이어트 식단이 아닌 식단의 일부이다. (표 3)에서는 1인분 기준 칼로리, 포함 성분 등이 포함되어 있어 다이어트의 적합한지를 판별한다.

(표 3) User1 제외 식단의 일부

112	제육덮밥(782.38kcal) - 탄수화물 : 115.48g - 단백질 : 30.08g - 지방 : 22.24g - 당류 : 7.27g - 나트륨 : 1538.06mg - 콜레스테롤 : 83.77mg - 포화지방산 : 5.68g - 트랜스지방 : 0.15g
2	치즈버거(454.5kcal) - 탄수화물 : 44.85g - 단백질 : 22.65g - 지방 : 20.85g - 나트륨 : 883.5mg
551	부대찌개(520.05kcal) - 탄수화물 : 47.34g - 단백질 : 25.96g - 지방 : 3.87g - 나트륨 : 2664.21mg - 콜레스테롤 : 82.64mg - 포화지방산 : 6.78g - 트랜스지방 : 0.11g

(표 3)의 식단과 같이 다이어트 식단이 아닌 식품들을 제외하며 사용자에게 적합한 다이어트 식품 중 Top-3 List를 구성하며 식단을 추천한다. (그림 7)은 기존의 방법으로 추출했을 때와 식단 분석을 통해 다이어트 식단 추출 후 추천한 방식을 평가하기 위해 facebook, instagram 설문 조사 기능을 통해서 사용자 97명에게 만족도를 검사를 진행한 결과이며 다이어트 식단 추출 후 추천한 방식이 사용자 만족도가 월등히 높은 것을 보여준다.



(그림 7) 사용자 만족도 조사

R은 무료소프트웨어재단의 GNU/GPL을 따르는 소프트웨어로 자유롭게 사용할 수 있으며, 다양한 통계분석 library와 강력한 그래픽 기능을 가지고 있다[7]. 때문에 본 논문에서는 R에서 제공하고 있는 'recommenderlab'를 이용하여 Jaccard Similarity를 이용한다. 이를 통해 사용자와 기존의 사용자 그룹사이의 유사도를 측정했고 협업 필터링을 이용해 유사 그룹을 형성한다. (그림 8)는 Jaccard Similarity 공식이다.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

(그림 8) Jaccard Similarity 공식

Jaccard Similarity 공식은 사용자가 1일 때 유사도를 측정하고 1과 근접 할수록 사용자와 유사도가 높다고 판단한다. 이러한 방법으로 유사도를 측정해 얻은 상위 그룹을 Top-N List로 구성했다.

이렇게 추천된 음식은 안드로이드 스튜디오에서 구현된 애플리케이션을 통해 편리하게 이용이 가능하며 식단의 사진과 항목을 확인 할 수 있다. (그림 9)는 User1과 User2의 개인화 요소이며 안드로이드 스튜디오에서 제공하는 에뮬레이터를 통해 추출한 user1의 식단 추천 결과 화면은 (그림 10)과 같다. 또한 (그림 10)의 식단은 (표 3)의 식단은 제외한 후 Top-3 List를 추출하여 추천하였다. user1의 식단과 user1과 같은 방식으로 추천 된 user2의 추천 식단은 (그림 11)와 같다.

ID	키	나이	체중	성별	추천식단
1	164	17	69	여	연어샐러드, 자몽, 아스파라거스
2	182	24	97	남	참치야채비빔밥, 단호박, 브리또

(그림 9) user1과 user2의 개인화 요소



(그림 10) user1의 추천 결과

ID	추천식단
1	아침 : 달걀 베이컨 샐러드 점심 : 오렌지, 오이, 당근, 블루베리 저녁 : 닭가슴살 덮밥
2	아침 : 야채비빔밥 점심 : 닭가슴살 샐러드 저녁 : 자몽

(그림 11) user1과 user2의 추천 식단

(그림 11)을 보면 (그림 9)의 user1과 user2의 추천 결과가 다른 것을 볼 수 있다. 이는 두 사용자의 개인화 요소가 다르며 추천 식단이 다르기 때문에 서로 다른 유사 그룹에 속했기 때문이다. 본 시스템은 이 처럼 개인에게 맞춤형 다이어트 식단을 제공하고 있다.

#### 4. 결론

본 논문에서는 식단 조절이 필요한 사용자에게 맞춤형 식단을 제공하기 위해 협업필터링을 이용하였고 사용자와 기존 사용자의 유사도를 측정하는데 사용되는 개인화 요인과 Open-API에서 추출한 식품 데이터베이스를 이용해

다이어트의 적합한 식단을 분석했고 이러한 추천 방식이 사용자 만족도가 더 높다는 것을 보였다. R을 통해 유사도 측정을 했으며 안드로이드 애플리케이션을 통해 보다 편리하게 다이어트 식단을 추천 받을 수 있도록 하였다. 향후 연구 과제는 기존의 데이터가 별로 없을 때 발생하는 Cold start 문제를 해결하여 더욱 정확한 추천이 가능할 것 이다.

#### 참고문헌

- [1] “국민건강영양조사”, 보건복지부 - mw.nhi.go.kr,
- [2] “서울시민 비만을 증가”, 연합뉴스 - <http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=102&oid=001&aid=0006338689>
- [3] “네이버 지식백과 비만”, 네이버 - <http://terms.naver.com/entry.nhn?docId=926838&cid=51007&categoryId=51007>
- [4] 홍성은, 김화중, 안두현 “사용자의 목적을 반영할 수 있는 음식 추천 시스템의 제안”, 한국통신학회, 2016
- [5] Sung-Mahn Ahn, In-Hwan Kim “Understanding the Performance of Collaborative Filtering Recommendation through Social Network Analysis”, Society for e-Business Studies, 2012
- [6] “네이버 지식백과 비만도 진단법”, 네이버 - <http://terms.naver.com>
- [7] 고영준, 박용민, 김진석 “WebER:R을 이용한 웹 기반의 교육용 통계 분석 시스템 구현”, 한국통계학회논문집, 19권, 2호, pp. 257-266, 2012