R에서 협업필터링과 개인화 요소를 이용한 개인 맞춤형 운동 추천 시스템

이한나*, 백수빈*, 박두순*
순천향대학교 컴퓨터소프트웨어공학과*
e-mail : hnjh02@naver.com, coffees1234@naver.com

Exercise Recommendation System using Collaborative Filtering and Personal Propensity in R System

Han-Na Lee*, Su-Bin Baeck*, Doo-Soon Park*
*Dept. of Computer Software Engineering, SoonchunHyang University

요 약

최근 자기관리의 필요성이 증가하면서 자신이 만족하는 신체를 갖기 위해 다이어트를 하는 사람들이 많아졌다. 다이어트를 하는 데 여러 방법이 존재하지만 건강한 다이어트를 위해서는 운동이 필수적으로 겸해져야 한다. 그러나 많은 사람들이 자신의 체형에 적합한 운동을 찾는 데 어려움을 겪는다. 따라서 개인의 체형과 성향에 맞는 운동을 추천해 줄 수 있는 추천 시스템이 필요하다.

본 논문에서는 나이, 신장, 몸무게, 체지방률, 식습관과 같은 개인화 요소와 협업 필터링 방법을 R 시스템을 사용하여 사용자 맞춤형 운동을 추천하는 추천 시스템을 제안한다.

1. 서론

최근 질적인 삶의 중요성과 건강에 대한이 필요성이 향상되면서 운동의 중요성에 대한 국민의 인지도는 점차적으로 높아지고 있다[1]. 운동은 인체생리학적으로 보면 심폐지구력을 증진시키고, 혈액순환을 도와주며, 다양한 호르몬과 효소의 활성을 증진시키는데 효과를 준다[2]. 운동은 우리 몸에 다양한 영향을 미친다. 그 중 운동이 질환에 미치는 긍정적인 영향에 대한 것은 (그림 1)과 같다[3].

(그림 1) 운동이 질환에 미치는 긍정적인 영향[3]

운동이 주는 많은 긍정적인 영향들로 인해 최근 건강관리에 대한 사람들의 관심이 많이 증가하였으나 정작 운동을 실천하는 사람들의 수는 그 수의 절반뿐이다. 시장조사전문기업 마크로밀 엠브레인의 트렌드 모니터가 전국 만 19세~59세 성인남녀 1,000명을 대상으로 '건강 및 외모 관리'에 대한 설문조사 실시 결과, 전체 응답자의 83%가 평소 건강관리가 필요한 편이라고 생각하는 것으로 나타났다[4]. 약 80%의 사람들이 건강관리의 필요성을 느끼고 있음에도 불구하고 정작 건강관리를 위해 노력하는 사람은 약 40%에 불과하다[4]. 최근 들어 운동의 안전과 효과의 측면이 강조되면서 개인별 운동능력과 선호도에 따라 운동처방을 받고 운동요법을 적용하는 프로그램이 시도되고 있으나 실제로 처방 받은 운동을 실행하도록 시작하고 유지, 지속하는 것이 쉽지 않다는 문제가 제기되고 있다[1]. 운동에 관한 전문적인 지식이 부족한 사람들은 헬스장에서 전문적인 트레이닝을 받지 않는다면 자신의 체형에 적합한 운동을 찾는 것에 어려움을 느낀다.

현재 출시되어 있는 어플리케이션들은 자신의 체형이나 환경에 따라 운동을 추천해주는 것이 아니라 단순히 운동 루틴과 반복시간 등을 추천해주는 것이 전부이기 때문에 자신의 체형에 알맞은 운동을 찾는 것에 대한 한계가 존재한다. 따라서 본 논문에서는 사람들에게 개인화된 운동 방법을 제공하기 위해서 사용자의 신체 정보에 관한 개인화 요소를 활용하여 협업 필터링과 k-최근접 이웃 알고리

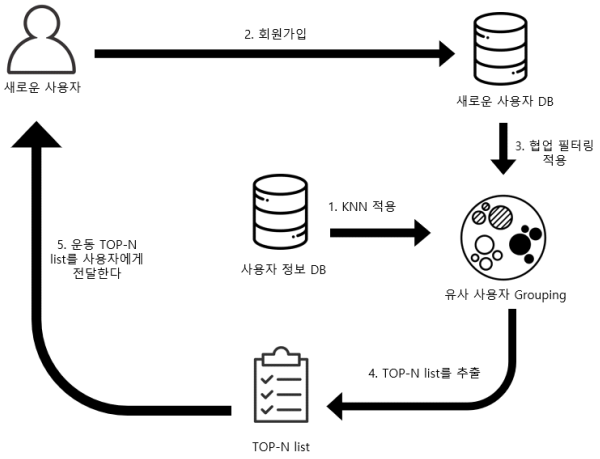
*corresponding author : 박두순
- 이 논문은 한국연구재단의 지원을 받아 수행되었음(No. NRF-2017R1A2B1008421)

즘을 이용해 사용자의 신체조건에 적합한 운동 방법을 추천하는 시스템을 제안한다.

2. 시스템의 구성

본 논문에서는 추천을 제공하기 위해서 협업필터링과 k-최근접 이웃 알고리즘을 사용한다. 패턴 인식에서, k-최근접 이웃 알고리즘은 분류나 회기에 사용되는 비모수 방식이다[5]. 협업 필터링(collaborative filtering)은 많은 사용자들로부터 얻은 기호정보에 따라 사용자들의 관심사들을 자동적으로 예측하게 해주는 방법이다[6]. 협업 필터링 접근법의 근본적인 가정은 사용자들의 과거의 경향이 미래에서도 그대로 유지 될 것이라는 전제에 있다[6].

본 논문에서 구현할 추천 시스템 시나리오는 (그림 2)와 같다.



(그림 2) 운동 추천 시스템 시나리오

(그림 2)의 시나리오에 대한 설명은 아래와 같다.

(1) 5가지의 개인화 요소 (나이, 신장, 몸무게, 체지방량, 식습관) 가 포함되어있는 기존 사용자 정보를 서버에 전달한다.

(2) 서버에 전달한 기존 사용자의 정보를 사용자 정보 데이터베이스에 저장한다.

(3) 데이터베이스에 저장된 사용자의 정보를 유클리디안 거리 공식을 사용해 사용자 간 유사도를 비교한 후 유사 사용자끼리 knn알고리즘을 사용해 Grouping한다.

(4) 새로운 사용자가 회원가입을 하면서 자신의 체형 정보를 입력하게 하고 입력받은 체형 정보를 새로운 사용자의 데이터베이스에 저장한다.

(5) 새로운 사용자가 서버에 운동 추천을 요청한다.

(6) 새로운 사용자 데이터베이스와 Grouping된 유사 사용자의 데이터베이스와 협업필터링을 적용하여 새로운 사용자와 가장 유사한 그룹을 찾는다.

(7) 구해진 가장 유사한 그룹 내에서 TOP-N 유저를 추출하여 생성한 TOP-N list를 새로운 사용자에게 추천해준다. 운동을 도출해 TOP-N list를 생성한다.

(8) 생성된 TOP-N list를 서버에 전달한다.

(9) 서버로 전달된 추천 운동 목록을 사용자에게 전달하여 운동을 추천해준다.

본 논문에서는 5가지의 개인화 요소를 수치화하여 Knn알고리즘과 협업필터링을 적용한다. 그 수치화된 개인화 요소는 (표 1)과 같다.

(표 1) 개인화 요소 수치화한 표

A	B	C	D	E	F
user_id	age	weight	tall	BodyFat	eating
1	3	3	3	4	3
2	4	2	6	6	4
3	3	2	5	1	3
4	3	2	5	3	1
5	3	2	7	3	4
6	3	2	6	3	3
7	2	2	6	4	2
8	1	2	4	5	3
9	2	2	4	4	1
10	2	2	5	6	2
11	1	2	6	1	1
12	2	2	5	6	1
13	1	2	4	7	3
14	3	2	3	4	1
15	3	7	8	6	1
16	3	2	7	5	3
17	2	2	7	4	3
18	1	2	3	5	3
19	2	2	3	3	2
20	3	2	6	5	3
21	3	2	6	4	2
22	2	2	4	3	3
23	3	2	3	3	1
24	2	2	4	5	3
25	5	8	5	7	2
26	2	2	5	4	2
27	2	2	4	10	2
28	5	2	6	5	3
29	1	2	3	7	2
30	2	2	6	2	1

3. 시스템의 구현

본 논문에서는 데이터 구축을 위해 kaggle에서 수집한 운동 설문 조사 데이터[8]를 토대로 운동과의 상관관계가 높다고 생각되는 5가지(나이, 키, 몸무게, 체지방량, 식습관)의 개인화 요소를 선정해 사용하였고, 사용자의 개인화 요소를 본 데이터에서 구분되어 있던 기준을 바탕으로 수치화해서 나타내었다. 또한 총 550개의 데이터를 사용하였고 500개의 데이터는 train데이터로, 나머지 50개의 데이터는 test데이터로 사용해 모델 구현 후 성능을 판단하고자 하였다. 구현은 R stdio에서 시행되었다. (그림 3)은 R stdio에서 읽어드린 수치화된 사용자 개인화 요소 데이터이다.

	age	weight	tall	BodyFat	eating
1	3	3	3	4	3
2	4	2	6	6	4
3	3	2	5	1	3
4	3	2	5	3	1
5	3	2	7	3	4
6	3	2	6	3	3
7	2	2	6	4	2
8	1	2	4	5	3
9	2	2	4	4	1
10	2	2	5	6	2
11	1	2	6	1	1
12	2	2	5	6	1
13	1	2	4	7	3
14	3	2	3	4	1
15	3	7	8	6	1
16	3	2	7	5	3
17	2	2	7	4	3
18	1	2	3	5	3

(그림3) R studio에서 읽은 사용자 개인화 요소 데이터

원래의 설문조사 데이터를 토대로 나이는 18살 미만(1), 18살~21살(2), 22살~26살(3), 27살~30살(4), 31살 이상(5)로 5가지로 수치화를 진행하였다. 몸무게는 59kg 이하(1), 60kg~66kg(2), 67kg~74kg(3), 75kg~83kg(4), 84kg~93kg(5), 94kg~105kg(6), 106kg~120kg(7), 121kg 이상(8)으로 8가지로 수치화를 진행하였다. 키는 152 미만(1), 152cm~165cm(2), 166cm~170cm(3), 171cm~175cm(4), 176cm~180cm(5), 181cm~185cm(6), 186cm~191cm(7), 192cm 이상(8)으로 8가지로 수치화를 진행하였다. 체지방률은 10% 미만(1), 11%~12%(2), 13%~14%(3), 15%~16%(4), 17%~18%(5), 19%~20%(6), 21%~22%(7), 23%~24%(8), 25%~26(%)(9), 27% 이상(10)으로 10가지로 수치화를 진행하였다. 식습관은 생식 위주 식단(1), 자신의 체형을 고려한 식단(2), 체중조절식단(3), 비식이요법(4)으로 4가지로 수치화를 진행하였다.

이러한 방식으로 수치화한 데이터로 사용자 간 유사도 거리를 측정해 유사한 사용자를 식별하고 유사한 사용자끼리 Knn 방식을 사용해 그룹핑을 한다. 이때 사용되는 유사도 공식은 Knn 방식의 default값으로 지정되어있는 유클리디안 거리 공식을 사용한다. 유클리디안 거리 공식은 다음 (식1)과 같다.

$$d(A,B)=\sqrt{\sum_{i=1}^{n}(A-B)^2}$$

(식1) 유클리디안 거리 공식

이 공식은 두 사용자 위치 사이의 유클리드 거리를 계산하고 거리가 멀면 사용자 간의 유사도가 낮고 거리가 가까우면 사용자 간의 유사도가 높다는 것을 의미한다. 유클리디안 거리 공식을 사용해 사용자 간 유사도를 구한 후 Knn(K nearest neighbor)알고리즘을 사용해 유사한 사용자끼리 총 100개의 그룹으로 그룹핑을 했을 때 (그림4)와 같았다.

	age	weight	tall	BodyFat	eating	group.cluster
1	3	3	3	4	3	46
2	4	2	6	6	4	53
3	3	2	5	1	3	49
4	3	2	5	3	1	34
5	3	2	7	3	4	67
6	3	2	6	3	3	49
7	2	2	6	4	2	85
8	1	2	4	5	3	24
9	2	2	4	4	1	36
10	2	2	5	6	2	73
11	1	2	6	1	1	29
12	2	2	5	6	1	78
13	1	2	4	7	3	8
14	3	2	3	4	1	62
15	3	7	8	6	1	57

(그림4) 100개의 그룹으로 그룹핑 된 사용자

그룹핑을 마친 후 새로운 사용자가 자신의 개인화 요소를 입력하고 운동 추천 요청을 하게 되면 협업필터링을 통해 가장 유사한 그룹에 포함되어있는 사람들로 TOP-N 유저를 추출하여 생성한 TOP-N list를 새로운 사용자에게 추천해준다.
(그림 5)는 501, 502, 503 사용자의 개인화 요소이다.

A	B	C	D	E	F
user_id	age	weight	tall	BodyFat	eating
501	3	2	6	5	3
502	2	2	6	4	3
503	2	2	5	4	3

(그림 5) 501, 502, 503 사용자의 개인화 요소

유사한 그룹에 있는 사용자에게 추천해 준 추천 결과에 따라 Top3-List를 추출해주었고 각 사용자 별로 추출된 결과는 (그림6), (그림7), (그림8) 과 같다.

	501
1	General Strength
2	Powerlifting
3	Hypertrophy

(그림6) 501 사용자 Top3-List

(그림6)은 age(3), weight(2), tall(6), BodyFat(5), eating(3)의 개인화 요소를 가지고 있는 501번 사용자에게 추천해 준 Top3-List이다. 501번 사용자에게는 첫 번째로 근력강화 운동, 두 번째로 리프팅 운동, 세 번째로 유산소 운동이 추천되었다.

	502
1	General Strength
2	Hypertrophy
3	Olympic Weightlifting Routines

(그림7) 502 사용자 Top3-List

(그림7)은 age(2), weight(2), tall(5), BodyFat(4), eating(3)의 개인화 요소를 가지고 있는 502번 사용자에게 추천해 준 Top3-List이다. 502번 사용자에게는 첫 번째로 근력강화 운동, 두 번째로 유산소 운동, 세 번째로 운동 선수 루틴이 추천되었다.

	503
1	General Strength
2	Hypertrophy
3	Olympic Weightlifting Routiones

(그림8) 503 사용자 Top3-List

(그림8)는 age(2), weight(2), tall(5), BodyFat(4), eating(3)의 개인화 요소를 가지고 있는 503번 사용자에게 추천해 준 Top3-List이다. 503번 사용자에게는 첫 번째로 근력강화 운동, 두 번째로 유산소 운동, 세 번째로 운동 선수 루틴이 추천되었다.
결론적으로 tall을 제외한 나머지 성향이 같은 502번 사용자와 503번 사용자는 유사도가 높아 같은 운동이 추천된 것을 볼 수 있고 각 사용자를 비교해보았을 때 5개의 개인화 요소 중 키보다는 나이와 체지방량이 운동과 더 높은 상관관계를 가지고 있음을 알 수 있다.

4. 결론

본 논문에서는 정확성을 높이기 위해 5가지의 개인화 요소를 사용하였고 유클리디안 거리 공식을 사용해 사용자 간 유사도를 측정해 Knn을 이용하여 그룹핑한 유사사용자와 신규 사용자의 협업 필터링을 통해 신규 사용자에게 운동을 추천해주는 시스템을 구현하였다. 기존 추천시스템에서는 개인화 요소가 부족하여 정확성이 다소 떨어지는 점을 보완하고자 운동과의 상관관계가 있는 요소들을 최대한 사용하여 구현하였다.

향후 연구 과제는 설문 조사 데이터가 아닌 더 신뢰성이 확실한 데이터를 사용해 시스템의 성능을 높이고, 유사도를 구하는 다른 여러 알고리즘을 사용하여 각각의 성능을 비교해 가장 좋은 성능을 가진 알고리즘으로 추천 시스템을 구현할 수 있도록 할 것이다.

참고문헌

[1] 황애린, 유지수, 김춘자, “계획된 운동프로그램이 제2형 당뇨병 환자의 대사, 심폐기능 및 운동이행에 미치는 영향,” - 대학간호회지, 제31권, 제1호, pp. 20-30, 2000

[2]https://m.lawtimes.co.kr/Content/Opinion?serial=146231

[3]http://health.chosun.com/site/data/html_dir/2018/03/07/2018030700852.html

[4] http://eknews.net/xe/kr_politics/489476, 2016. 11 춘계학술발표대회

[5] https://ko.wikipedia.org/wiki/K-%EC%B5%9C%EA%B7%BC%EC%A0%91_%EC%9D%B4%EC%9B%83_%EC%95%8C%EA%B3%A0%EB%A6%AC%EC%A6%98

[6] https://ko.wikipedia.org/wiki/%ED%98%91%EC%97%85_%ED%95%84%ED%84%B0%EB%A7%81

[7] https://www.kaggle.com/raford2/2017-census-data-for-4chans-fitness-board - 4chan의 피트니스 보드에 대한 2017 인구 조사 데이터