

R에서 협업필터링과 K-NN을 이용한 개인 맞춤형 운동 추천 시스템

백수빈¹, 박두순²

¹순천향대학교 컴퓨터소프트웨어공학과

coffees1234@naver.com¹

Personalized Exercise Recommendation System using Collaborative Filtering and K-NN in R System

Su-Bin Baeck¹, Doo-Soon Park¹⁾²

¹⁾²Dept. of Computer Software Engineering, Soonchunghyang University

요 약

최근 질적인 삶의 중요성과 건강에 대한 필요성이 향상되면서 운동의 중요성에 대한 국민의 인지도가 증가했다. 체력적인 효과 심리적인 효과 면역효과 등 운동이 주는 많은 긍정적인 영향들로 인해 최근 건강관리에 대해 사람들의 관심이 많이 증가했으나 자신에게 알맞는 운동 방법을 알지 못해 정작 운동을 실천하는 수는 그 수의 절반뿐이다. 따라서 개인의 신체 알맞는 운동을 추천해 줄 수 있는 추천 시스템이 필요하다. 본 논문에서는 신장, 몸무게, 나이, 주당 운동 횟수, 성별과 같은 개인화 요소를 이용한 협업 필터링과 k-nn 을 R 시스템을 사용하여 사용자 개인 맞춤형 운동 추천 시스템을 제안한다.

1. 서론

최근 질적인 삶의 중요성과 건강에 대한 필요성이 향상되면서 운동의 중요성에 대한 국민의 인지도는 점차적으로 높아지고 있다[1]. 운동은 세포 자체의 활동성을 높여서 심장, 폐, 혈관, 근육 등 여러 종류의 세포로 이루어진 인체 기관인 형태와 기능을 발달하게 해주며, 생리적 노쇠현상을 지연시키는 효과를 가지고 있다[2]. WHO에 따르면 신체활동은 식이조절과 함께 과체중 및 비만을 감소시키고, 각종 암, 특히 최근에 증가하고 있는 대장암과 유방암을 예방하는 효과가 있다고 밝혔다[3]. 또한, 최근에 발표된 한국보건사회연구원 보고에 따르면, 신체 활동이 치매발병 위험요인을 가장 효과적으로 낮추는 것

으로 나타났다[3].

운동이 주는 많은 긍정적인 영향들로 인해 최근 건강관리에 대한 사람들의 관심이 많이 증가하였으나 정작 운동을 실천하는 사람들의 수는 그 수의 절반도 되지 않는다. 2016년 국민건강영양조사(질병관리본부) 결과에 따르면, 우리나라 국민의 걷기실천율은 39.6%, 유산소 신체활동 실천율은 49.4%, 유산소 및 근력운동 실천율은 15.0%인 것으로 나타났다[3]. 이중 신체활동 실천여부를 판단할 수 있는 대표지표인 ‘유산소 및 근력운동 실천’은 남자가 5명중 1명, 여자는 10명 중 1명에 그치는 상황이다[3]. 최근 들어 운동의 안전과 효과의 측면이 강조되면서 개인별 운동능력과 선호도에 따라 운동 처방을 받고 운동요법을 적용하는 프로그램이 시도되고 있으나 실제로 처방 받은 운동을 실행하도록 시작하고 유지, 지속하는 것이 쉽지 않다는 문제가 제기되고 있다[1]. 운동하는 방법에 대한 전문적인 지식이 부족한 사람들은 트레이너의 전문적인 지도를 받지 않는다면 자신의 신체에 알맞은 운동을 하는데 어려움을 느낀다.

현재 출시되어있는 어플들은 단순히 운동 루틴과 스

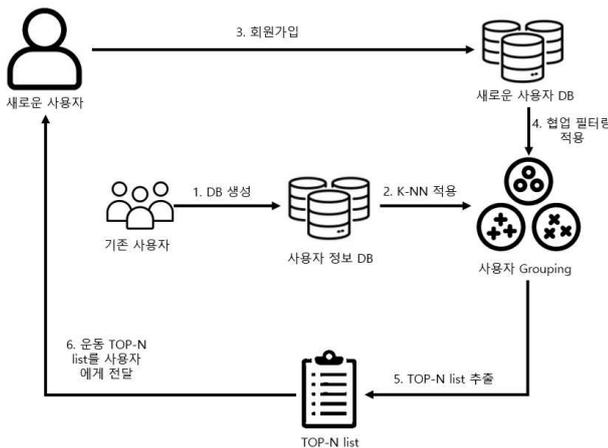
corresponding author: Doo-Soon Park

-Acknowledgements: This research was supported by the National Research Foundation of Korea (No. NRF-2022R1A2C1005921) and BK21 FOUR (Fostering Outstanding Universities for Research) (No.5199990914048) and the MSIT(Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation) in 2021”(2021-0-01399)

케줄링만을 추천해주고 있어 자신의 체형에 알맞은 운동을 찾는 것에 한계가 존재한다. 따라서 본 논문에서는 개인 맞춤형 운동을 제공하기 위해서 운동과 상관관계가 있는 사용자의 신체 정보에 관한 개인화 요소를 통해 협업 필터링과 k-최근접 이웃 알고리즘을 이용해 사용자의 신체조건에 알맞은 운동 방법을 추천하는 시스템을 제안한다.

2. 시스템의 구성

본 논문에서는 추천을 제공하기 위해서 협업 필터링과 k-최근접 이웃 알고리즘을 사용한다. 협업 필터링(collaborative filtering)은 많은 사용자들로부터 얻은 기호정보에 따라 사용자들의 관심사들을 자동적으로 예측하게 해주는 방법이다[4]. 패턴 인식에서, k-최근접 이웃 알고리즘은 분류나 회기에 사용되는 비모수 방식이다[5].



(그림 1) 운동 추천 시스템 시나리오

(그림 1)의 시나리오에 대한 설명은 아래와 같다.

- (1) 5가지의 개인화 요소 (나이, 신장, 몸무게, 주당 운동횟수, 성별)가 포함되어있는 기존 사용자 정보를 서버에 전달한다.
- (2) 서버에 전달한 기존 사용자의 정보로 사용자 정보 데이터베이스를 생성한다.
- (3) 데이터베이스에 저장된 기존 사용자의 정보를 유클리디안 거리 공식을 사용해 사용자 간 유사도를 비교한 후 유사 사용자끼리 k-nn알고리즘을 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 사용자의 데이터베이스에 협업필터링을 적용하여 새로운 사용자와 가장 유사한 그룹을 찾는다.
- (7) 구해진 가장 유사한 그룹 내에서 최대 개수 counting을 이용해 TOP-N list를 생성한다.
- (8) 생성된 TOP-N list를 서버에 전달한다.
- (9) 서버로 전달된 추천 운동 목록을 새로운 사용자에게 전달하여 운동을 추천해준다.

본 논문에서는 5가지의 개인화 요소를 수치화하여 k-nn알고리즘과 협업필터링을 적용한다. 그 수치화된 개인화 요소는 <표 1>과 같다.

<표 1> 5가지 개인화 요소 수치화한 표

	A	B	C	D	E
1	성별	나이	신장	주당 운동횟수	몸무게
2	1	1	4	3	6
3	1	1	4	3	5
4	1	2	4	2	7
5	1	2	5	2	5
6	1	2	4	2	6
7	1	2	4	3	8
8	2	1	3	1	2
9	1	1	4	2	5
10	2	2	3	1	2
11	2	2	3	2	4
12	2	2	2	1	3
13	2	2	3	1	2
14	1	1	5	3	7
15	2	2	3	1	1
16	2	3	2	2	2
17	1	3	3	4	5
18	2	3	3	2	2
19	1	1	5	5	6
20	2	1	3	3	3
21	1	2	4	2	6
22	1	2	4	3	8
23	2	1	3	1	2
24	1	1	4	2	5
25	2	2	3	1	2
26	1	3	3	2	4
27	2	2	2	1	3
28	2	1	3	1	2
29	1	1	5	3	7
30	2	2	3	1	1
31	2	3	2	2	2

3. 시스템의 구현

본 논문에서는 데이터 구축을 위해 설문조사를 하여 수집한 운동 설문 조사 데이터를 토대로 운동과의 상관관계가 높다고 생각되는 5가지 개인화 요소(나이, 신장, 몸무게, 주당 운동횟수, 성별)를 선정해 사용하였고, 사용자의 개인화 요소를 본 데이터에서 구분되어 있던 기준을 바탕으로 수치화해서 나타내었다. 또한, 총 100개의 데이터를 사용하였고 50개의 데이터는 학습데이터로, 나머지 50개의 데이터는 테스트 데이터로 사용해 모델 구현 후 성능을 판단하고자 하였다. 구현은 R studio에서 시행되었다. 개인화 요소를 수치화한 <표1>의 데이터로 사용자 간 유사도 거리를 측정해 유사한 사용자를 식별하고 K-nn 방식을 사용해 유사한 사용자끼리 Grouping을 한다. 이때 사용되는 유사도 공식은 K-nn 방식

의 default값으로 지정되어있는 유클리디안 거리 공식을 사용한다. 유클리드 거리 공식은 다음 (식1)과 같다.

$$\|p - q\| = \sqrt{(p - q) \cdot (p - q)} = \sqrt{\|p\|^2 + \|q\|^2 - 2p \cdot q}$$

(식1) 유클리드 거리 공식

유클리드 거리란 두 점 사이의 거리를 계산할 때 쓰는 방법이다[6]. 즉, 이 공식은 두 사용자 위치 사이의 거리를 계산하는 공식이며, 거리가 멀면 사용자 간의 유사도가 낮고 거리가 가까우면 사용자 간의 유사도가 높다는 것을 의미한다.

이 공식을 사용해 사용자 간의 거리로 사용자 간의 유사도를 구한 후 K-nn알고리즘을 사용해 유사한 사용자끼리 Grouping을 한 결과는 (그림 2)와 같다.

	gender	age	traincont	weigh	height
1	1	1	3	5	4
2	1	1	2	5	4
3	1	1	2	6	5
4	1	2	3	6	5
5	1	2	3	6	6
6	1	2	4	5	5
7	1	2	4	5	5
8	1	2	2	6	5
9	1	2	2	6	5
10	1	2	2	7	6
11	2	2	2	7	7
12	2	2	3	7	7
13	2	2	3	8	5
14	1	2	4	6	5
15	1	2	3	6	6
16	1	2	3	6	6
17	1	2	3	5	6
18	1	3	3	6	5
19	1	3	3	7	5

(그림 2) Grouping된 유사 사용자 데이터

새로운 사용자가 서버에 운동 추천을 요청하면 새로운 사용자 데이터베이스와 Grouping된 유사 사용자의 데이터베이스에 협업필터링을 적용하여 새로운 사용자와 가장 유사한 그룹의 최대 개수 counting을 통해 운동 TOP-N list를 생성한다. 생성된 운동 TOP-N list를 사용자에게 추천해준다.

사용자 아이디 50, 51에게 추천된 유사한 그룹에 있는 사용자들의 운동 TOP-N list는 (그림 3)과 같다.

50	dead lif	squat	bench pr	barbell lursit up	lunge
51	dead lif	bench p	sit up	barbell lurussiantw	lunge

(그림 3) 사용자 추천 운동 TOP-N list

4. 결론

본 논문에서는 5가지의 개인화 요소를 사용하였고 유클리디안 거리 공식을 사용해 사용자 간의 유사도를 측정해 k-nn을 이용하여 Grouping한 유사 사용자와 새로운 사용자의 협업 필터링을 통해 새로운 사용자에게 운동을 추천해주는 시스템을 구현하였다. 기존 추천시스템에서는 개인화 요소가 부족하여 정확성이 다소 떨어지는 점을 보완하고자 운동과의 상관관계가 높은 요소들을 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 제 2형 당뇨병 환자의 대사, 심폐기능 및 운동이행에 미치는 영향,” - 대학간호회지, 제31권, 제1호, pp. 20-30, 2000
- [2]<https://www.ydp.go.kr/health/contents.do?key=3637&>
- [3]<http://www.healthinnews.co.kr/news/articleView.html?idxno=442>
- [4]https://ko.wikipedia.org/wiki/%ED%98%91%EC%97%85_%ED%95%84%ED%84%B0%EB%A7%81
- [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/%EC%9C%A0%ED%81%B4%EB%A6%AC%EB%93%9C_%EA%B1%B0%EB%A6%AC