빅 데이터 기반의 식습관 분석 및 관련 상품 추천 온라인 몰 API

장소은*, 김문현**,나지현**, 홍장의*
email: thdmsdia@gmail.com

*충북대학교 소프트웨어학과
**충북대학교 국제경영학과

Big Data based Diet Analysis and Relevant Product Recommendation Online-mall API

Soe-Un Jang*, Moon-Hyun Kim**, Ji-Hyun Na**, Jang-Eui Hong*

*Dept of Computer Science, Chungbuk National University
**Dept of International Business, Chungbuk National University

요 약

최근 현대인들은 식습관이 불규칙하고 서구화되면서, 건강상의 많은 문제를 겪고 있다. 이와 더불어 1인 가구의 증가와 간단한 구매 방법 등으로 인해 온라인 몰 사용자가 늘어나고 있다. 본 프로젝트는 이러한 추세를 바탕으로, 사용자가 자주 사용하는 온라인 몰에 축적된 데이터를 기반으로 사용자의 식습관을 분석한다. 뿐만 아니라, 이를 바탕으로 구매 패턴을 분석하여 사용자의 영양 상태를 개선시킬 수 있는 상품 추천 서비스를 제공한다. 사용자는 자주 사용하는 온라인 쇼핑몰에서 상품 구매를 함과 동시에 구매한 상품에 대해 시각화된 영양소 분석 결과와 구매 패턴 분석 결과를 제공받을 수 있다. 본 논문에서는 개발한 API를 통해 사용자는 부족한 영양소를 쉽게 파악하여 효율적으로 건강관리를 할 수 있게 된다. 더 나아가, 자신의 구매 패턴을 파악할 수 있게 되어 현명한 소비 습관을 만드는 데에 기여할 수 있다.

1. 서론

최근 현대인들은 식습관이 불규칙하고 서구화되면서 많은 건강상의 문제를 겪고 있다. 이러한 변화로 인해 과거 성인에게만 발생하던 병들이 최근 젊은이들에게도 적지 않게 발생하고 있어 노소를 불문하고 건강관리에 대한 인식이 변화하고 있다. 또한 경제 성장에 따라 건강관리의 중요성이 높아지고 있음에 따라, 웰빙, 다이어트 관련 신산업의 규모가 나날이 성장하고 있다[1].

이러한 변화는 IT, 컴퓨터 시장에도 영향을 미치고 있다. 그 예로 여러 다이어트 애플리케이션 및 웹이 등장하고 있고 그중 몇몇 서비스들은 많은 회원 수를 확보하고 있으며 그중 일부는 유료로 사용할 정도로 인기를 끌고 있다[2]. 하지만 이러한 서비스들의 단점은 건강, 신체 관련 정보를 사용자가 일일이 수동으로 입력을 해야 분석 정보를 얻을 수 있다는 불편한 점이 있다[3]. 따라서 이 논문에서는 개인 정보를 수동으로 입력할 필요 없이, 개인의 식품 구매 기록을 데이터로 받아 이에 따른 소비 습관을 분석하여 개인의 식습관에 대한 분석 결과를 제공하기 위한 방법을 제시한다.

또한, 1인 가구 증가 및 식품 온라인 쇼핑몰의 신선도 유지 기술의 발달에 따라, 온라인 식품 쇼핑 사용자가 늘고 있다. 몰 회원들에 대한 구매 정보는 방대한 빅데이터를 형성하고 있으며, 이로부터 행사 기획, 고객 맞춤 서비스, 제품 전시 등과 같은 많은 비즈니스 활동에 활용도고 있다. 이와 함께, 고객을 위한 식습관 분석이나, 온라인 몰에서의 추천 시스템의 연구 또한 진행되고 있다[1-3].

본 논문에서 소개하는 API(Application Program Interface)는 이러한 시장 트랜드 변화에 따라, 사용자가 자주 사용하는 온라인 쇼핑몰에서 구매한 상품들의 패턴을 분석하여 구매 패턴과 관련된 새로운 상품을 추천해 준다. 본 논문에서 개발한 API를 통해 사용자는 부족한 영양소를 쉽게 파악하여 효율적으로 건강관리를 할 수 있게 된다. 더 나아가, 자신의 구매 패턴을 파악할 수 있게 되어 현명한 소비 습관을 만드는 데에 기여할 수 있다.

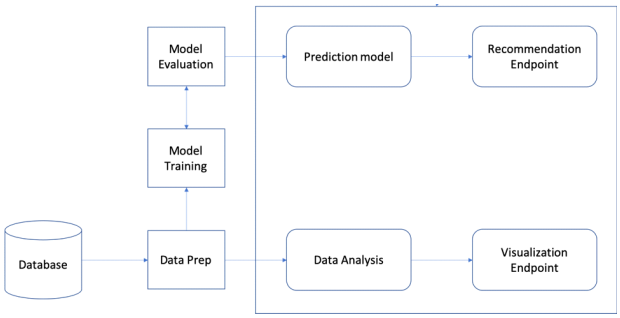
본 논문은 총 4장으로 구성되어 있다. 먼저, 1장에서는 본 논문과 관련된 최근 기술의 동향을 설명하고, 이러한 기술들의 문제점을 설명하여 이를 해결하기 위한 새로운 API를 제시한다. 2장에서는 제안하는 시스템에 대한 시스템 구조와 핵심 기능을 설명한다. 3장에서는 논문에서 제시한 API의 프로토타입을 제시하고, 이를 사례에 적용한 결과를 4장과 5장에 기술한다.

2. 식습관 분석 및 관련 상품 추천 시스템

2.1 제안 시스템 구조

기존의 빅 데이터 분석을 통한 추천 시스템은 이미 시장에 널리 상용화되고 있다. 그러나 상용화된 시스템들은 전반적인 상품에 관한 단순한 추천만을 제공할 뿐, 건강 증진과 관련된 정보 및 관련 상품 추천은 부족한 실정이다. 이러한 점을 보완하고, 웰빙 라이프를 추구하는 현재 시장의 트랜드에 발맞추기 위해 빅 데이터 분석을 통한 사용자 식습관 분석 및 이를 기반으로 한 상품 추천 시스템을 개발하고자 한다.

단순하게 사용자의 소비 패턴을 분석해 상품을 추천하는 기존 방식에서 더 나아가 사용자가 건강한 식습관을 형성할 수 있도록, 구매하려는 제품의 영양 정보를 시각화하여 제공한다. 이를 통해 사용자의 영양 소비 패턴을 분석하여 균형 잡힌 영양소 섭취 식단을 유지 및 관리할 수 있도록 도와주어 사용자의 건강 증진에 기여한다.

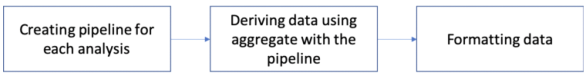


(그림 1) 시스템 구조도

그림 1은 앞에서 설명한 목적을 달성하기 위한 시스템의 구조를 나타낸 것으로 다음과 같은 컴포넌트들로 구성된다.

- 예측 모델(Prediction Model): 추천 시스템의 알고리즘이 구현된 부분으로, Associate rules 알고리즘에 식습관에 대한 평가 요소가 추가된다.
- 데이터 분석(Data Analysis): 프로세싱 된 주문 데이터를 기반으로 빅 데이터 분석을 통해 사용자의 식습관 분석을 위한 컴포넌트이다.
- 추천 엔드 포인트(Recommendation Endpoint): 본 발명의 추천 서비스를 API 형식으로 제공할 수 있도록 하는 Endpoint이다.
- 시각화 엔드 포인트(Visualization Endpoint): 본 발명의 사용자 식습관 분석의 결과를 시각적으로 보여주기 위한 서비스를 API 형식으로 제공할 수 있도록 하는 Endpoint이다.
- 이외에도 Backend로써, 데이터 기반의 학습과 학습 결과를 분석하기 위한 기능이 포함된다.

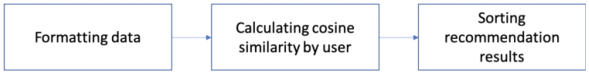
2.2 (핵심 기능) 사용자 식습관 분석 기능



(그림 2) 식습관 분석 로직

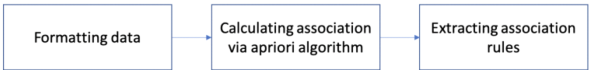
그림 2는 사용자의 식습관을 분석하기 위한 로직을 나타낸 그림이다. 먼저, 사용자 식습관에 맞는 파이프라인을 생성한다. 일반적으로 파이프라인이란 한 데이터 처리 단계의 출력이 다음 단계의 입력으로 이어지는 형태로 연결된 구조를 의미한다. 식습관 분석 로직에서 사용된 파이프라인은 데이터 파이프라인으로 보안성과 신뢰성을 갖춘 국제적인 공급망을 통해 데이터 교환에 대한 혁신적인 접근을 제공하는 것이다[4]. 그 후 해당 데이터 파이프라인을 통해 Aggregation을 사용해 분석 결과를 도출해낸다. 여기서 사용된 Aggregation은 데이터 프로세싱 파이프라인의 컨셉을 모델로 한 mongoDB의 Operation중 하나로서, 데이터를 프로세스하고 컴퓨팅 결과를 반환해주는 역할을 한다[5]. 이렇게 도출된 데이터를 프로세싱하여 사용자에게 식습관 분석 결과를 제공한다.

2.3 식습관 관련 상품 추천 기능



(그림 3) 식습관 관련 상품 추천 로직

그림 3은 식습관과 관련된 상품을 추천하기 위한 로직을 나타낸다. 식습관 관련 상품 추천에서 사용된 주요 알고리즘은 코사인 유사성(Cosine Similarity)이다. 코사인 유사성이란 추천 시스템에서의 필터링에서 사용되는 방식으로, 두 벡터의 코사인을 계산하여 두 벡터의 유사성을 도출해 낼 수 있다[6]. 본 기능에서는 데이터 프로세싱을 통하여 두 벡터를 생성하고 코사인을 계산한다. 코사인 결과 값을 통해 상품 간의 유사성을 도출한다. 이 코사인 유사성을 기반으로 사용자별 상품 추천 결과를 생성하여 정렬한다.



(그림 4) Association rule 도출 로직

그림 4는 Association rule을 도출하기 위한 로직을 설명한 그림이다. Association rule이란 data mining의 기법 중 데이터의 support, confidence 등을 계산하여 패턴을 발견한다[7]. 본 프로젝트에서 Association rule을 도출하기 위해 Apriori 알고리즘이 사용하였다.

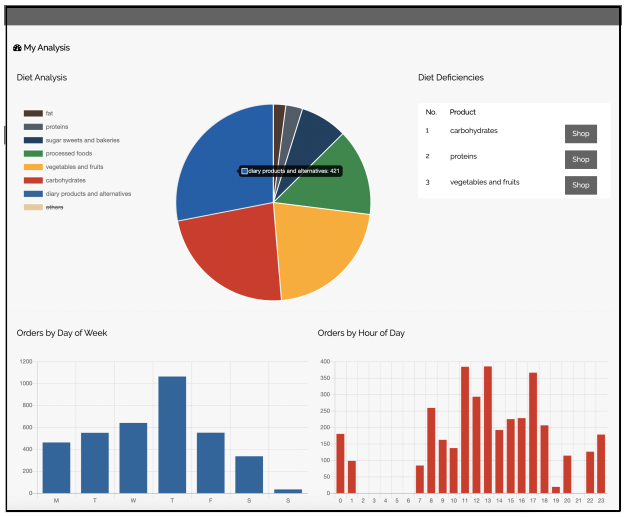
Apriori 알고리즘이란 Association rule을 사용하여 후보 아이템 리스트를 만들어낸다[8]. 본 프로젝트에서는 Apriori 알고리즘을 사용하기 위해, 구매 데이터를 그에 맞는 형식으로 프로세싱한다. 프로세싱 된 데이터를 이용하여 Apriori 알고리즘을 적용하여 Association rules를 도출한다.

3. 시스템 구현

사용자는 본인이 가지고 있는 ID/PW를 활용해 사이트에 시스템에 로그인 할 수 있다. 개인별 아이디에 할당된 User_id가 있고, 해당 소비자의 구매 내역은 mongo DB에 저장되어 있다. 소비자는 웹페이지에 로그인 후, 메뉴에 접근하여 식습관 분석이나 관련 상품 추천에 대한 정보를 열람할 수 있다.

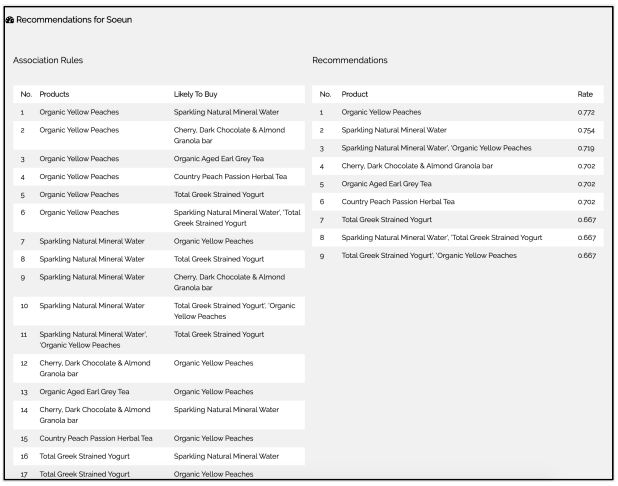
사용자는 구매내역에 따른 식습관 분석(Analysis) 메뉴와 분석 결과에 따른 상품 추천(Recommendation) 메뉴를 사용할 수 있다.

먼저, 식습관 분석 메뉴에서는 상품 구매내역 분석을 통한 영양소 분석 결과를 시각화된 그래프로 확인할 수 있다(그림 5). 그리고 이러한 영양소 분석을 기반으로, 부족한 영양소와 관련된 상품 추천 또한 확인할 수 있다. 각 영양소의 'shop' 버튼을 클릭하면, 사용자가 자주 사용하는 온라인 쇼핑몰로 이동하여 각각의 영양소를 보충할 수 있는 상품을 구매할 수 있다.



(그림 5) 식습관 분석 화면

다음으로 상품 추천 메뉴에서는, 사용자의 구매 기록을 분석하여 구매 패턴을 파악한다. 이를 통해, 사용자의 건강 증진을 위한 맞춤 식품 상품을 추천한다(그림 6의 오른쪽 목록). 이뿐만 아니라 이러한 상품 추천과 함께, 사용자가 함께 구매할만한 가능성이 높은 순으로 사용자에 연관 상품을 추천한다(그림 6의 왼쪽 목록).



Recommendations for Soeun

Association Rules

No.	Products	Likely To Buy
1	Organic Yellow Peaches	Sparkling Natural Mineral Water
2	Organic Yellow Peaches	Cherry, Dark Chocolate & Almond Granola bar
3	Organic Yellow Peaches	Organic Aged Earl Grey Tea
4	Organic Yellow Peaches	Country Peach Passion Herbal Tea
5	Organic Yellow Peaches	Total Greek Strained Yogurt
6	Organic Yellow Peaches	Sparkling Natural Mineral Water', 'Total Greek Strained Yogurt
7	Sparkling Natural Mineral Water	Organic Yellow Peaches
8	Sparkling Natural Mineral Water	Total Greek Strained Yogurt
9	Sparkling Natural Mineral Water	Cherry, Dark Chocolate & Almond Granola bar
10	Sparkling Natural Mineral Water	Total Greek Strained Yogurt', 'Organic Yellow Peaches
11	Sparkling Natural Mineral Water', 'Organic Yellow Peaches	Total Greek Strained Yogurt
12	Cherry, Dark Chocolate & Almond Granola bar	Organic Yellow Peaches
13	Organic Aged Earl Grey Tea	Organic Yellow Peaches
14	Cherry, Dark Chocolate & Almond Granola bar	Sparkling Natural Mineral Water
15	Country Peach Passion Herbal Tea	Organic Yellow Peaches
16	Total Greek Strained Yogurt	Sparkling Natural Mineral Water
17	Total Greek Strained Yogurt	Organic Yellow Peaches

Recommendations

No.	Product	Rate
1	Organic Yellow Peaches	0.772
2	Sparkling Natural Mineral Water	0.754
3	Sparkling Natural Mineral Water', 'Organic Yellow Peaches	0.719
4	Cherry, Dark Chocolate & Almond Granola bar	0.702
5	Organic Aged Earl Grey Tea	0.702
6	Country Peach Passion Herbal Tea	0.702
7	Total Greek Strained Yogurt	0.667
8	Sparkling Natural Mineral Water', 'Total Greek Strained Yogurt	0.667
9	Total Greek Strained Yogurt', 'Organic Yellow Peaches	0.667

(그림 6) 식습관 및 구매 패턴에 따른 상품 추천 화면

4. 사례적용

본 시나리오에서는 실제 사용자의 데이터가 아닌, 프로세싱 된 데이터를 사용하였다. 먼저 시나리오를 적용한 사용자의 과거 구매 목록은 그림 7과 같다.


```
"_id" : ObjectId("5cf941159297fa4f53d09653"), "" : 15520774, "order_id" : 1011604, "user_id" :
00330, "eval_set" : "prior", "order_number" : 2, "order_dow" : 1, "order_hour_of_day" : 13,
'days_since_prior_order" : 7, "product_id" : 47766, "add_to_cart_order" : 2, "reordered" : 1,
'product_name" : "Organic Avocado", "aisle_id" : 24, "department_id" : 4, "department" :
'produce", "aisle" : "fresh fruits", "foodgroup" : "vegetables and fruits" }
"_id" : ObjectId("5cf941159297fa4f53d09654"), "" : 15520775, "order_id" : 1011604, "user_id" :
00330, "eval_set" : "prior", "order_number" : 2, "order_dow" : 1, "order_hour_of_day" : 13,
'days_since_prior_order" : 7, "product_id" : 33719, "add_to_cart_order" : 3, "reordered" : 1,
'product_name" : "French Vanilla Yogurt Vanilla", "aisle_id" : 120, "department_id" : 16,
'department" : "dairy eggs", "aisle" : "yogurt", "foodgroup" : "diary products and alternatives" }
"_id" : ObjectId("5cf941159297fa4f53d09655"), "" : 15520776, "order_id" : 1011604, "user_id" :
00330, "eval_set" : "prior", "order_number" : 2, "order_dow" : 1, "order_hour_of_day" : 13,
'days_since_prior_order" : 7, "product_id" : 28613, "add_to_cart_order" : 4, "reordered" : 1,
'product_name" : "Organic Grapefruit Ginger Sparkling Yerba Mate", "aisle_id" : 94,
'department_id" : 7, "department" : "beverages", "aisle" : "tea", "foodgroup" : "others" }
```


(그림 7) 시나리오 사용자의 과거 구매 데이터

4.1 식습관 분석 사례 적용 시나리오

첫 번째로 사용자가 'Analysis' 메뉴를 클릭하면, 본인의 상품 구매 기록에 따른 영양소 분석 결과를 그래프로 확인할 수 있다. 사용자의 식습관 분석 결과는 그림 5와 같다.

먼저, 그림 5의 원형 그래프는 사용자의 상품 구매 분포를 나타낸다. 이를 분석해보면, 'Mini Original Babybel Cheese', 'Total Greek Strained Yogurt' 등과 같은 유제품 군을 가장 많이 구매하였다. 그 다음으로는 'Fruit & Nut Delight Bar', 'Cherry, Dark Chocolate & Almond Granola bar' 등과 같은 탄수화물 군, 'Organic Yellow Peaches', 'Dried Apples' 등의 과채류, 'Garlic & Parmesan Plentils', 'Everything Deli Style Pretzel Crisps Crackers' 등의 가공식품의 순서로 많이 구매한 것으로 나타났다. 이러한 구매 분포를 기반으로, 사용자에게 부족한 영양소 목록을 확인할 수 있다(그림 5의 오른쪽 목록). 이 목록에서 사용자는 가장 부족한 순서대로 3가지의 영양소를

파악할 수 있다. 사용자의 분석 결과에 따르면, 사용자는 현재 탄수화물, 단백질, 과채류 순으로 부족한 것으로 분석된다. 이에 따라, 사용자는 균형 있는 영양소 섭취를 위해, 각 영양소 이름 우측의 'Shop' 버튼을 클릭하면 사용자가 자주 이용하는 온라인 쇼핑몰로 이동하여, 해당 영양소와 관련된 상품들을 쇼핑할 수 있다. 이를 통해 사용자는 자신에게 부족한 각 영양소를 쉽게 파악하고 보충할 수 있게 된다.

4.2 상품 추천 사례 적용 시나리오

두 번째로, 사용자가 'Recommendation' 메뉴를 클릭하면, 사용자의 식습관과 구매 기록에 따른 구매 패턴 및 관련 상품 추천 결과를 확인할 수 있다(그림 6). 그림 6의 오른쪽의 'Recommendations' 목록에서 사용자는 'Organic Yellow Peaches'를 구매할 확률이 제일 높고, 그 다음으로는 'Sparkling Natural Mineral Water', 'Cherry, Dark Chocolate & Almond Granola bar' 등의 순으로 구매할 확률이 높은 것으로 분석된다.

왼쪽의 'Association Rules' 목록에서, 'Recommendations' 목록의 상품들에 대해 함께 구매할 확률이 높은 상품들을 확인할 수 있다. 해당 사용자는 1순위로 추천된 'Organic Yellow Peaches' 상품과 함께 'Organic Aged Earl Grey Tea', 'Country Peach Passion Herbal Tea' 그리고 'Total Greek Strained Yogurt' 등을 함께 구매할 것으로 예측된다.

4.3 적용 결과

이러한 시각화 자료 및 상품 추천 시스템은 기존의 다른 사이트의 구매내역만을 확인할 수 있는 것과는 다르게, 구매 내역 중 식품 영양성분과 관련된 정보를 분석한 결과들을 제공한다. 이 밖에도 데이터베이스에 저장된 여러 정보를 토대로 어떤 시간에 주로 어떤 주문을 했는지, 어떤 물건의 재주문율이 높은지에 대한 정보도 받아볼 수 있다.

분석 및 추천에 대한 응답시간은 2초 내외로 빠른 시간 내에 확인 할 수 있지만, 데이터베이스에 새로운 데이터들의 입출력이 없는 정적인 상태의 데이터를 사용한다는 점에서 실시간으로 데이터를 받는 시스템일 경우 응답 속도 측면에서 시간이 더 걸릴 수 있다는 제한점이 있다.

5. 결론 및 향후 연구

본 논문에서는 축적된 소비자들의 구매 내역을 바탕으로 식습관을 분석하고, 이를 통해 사용자별로 부족한 영양소를 보여주어 그와 관련된 상품을 추천받을 수 있도록 해주는 API를 제안한다. 나아가 소비자의 구매 내역을 통해 소비 습관을 파악하고, 사용자가 이후에 구매할 가능성이 높은 제품을 추천하여 소비자와 해당 업체 모두의 만족도를 높이는데 기여할 수 있다. 본 프로젝트에서는 데이터 셋을 외부에서 받아 사용하여 데이터의 추가적인 입, 출력이 없는 정적인 데이터 상태에서 개발을 진행하였다. 따라서 향후 연구에서는 실시간으로 구매 내역 정보를 받아오고 이러한 데이터를 기반으로 변화하는 소비자의 소비 패턴과 식습관 분석에 대해 학습을 진행하여, 소비자에게 더 나은 서비스를 제공할 수 있을 것이라 기대한다.

감사의 글

"이 연구는 과학기술정보통신부, 정보통신기획평가원의 SW 중심대학사업의 지원으로 수행되었음(2019-0-01183)"

참고문헌

[1] Sanjeevan Sivapalan, Alireza Sadeghian, Hossein Rahanam. "Recommender Systems in E-Commerce", August 2014, pp.2-5

[2] Frank B. Hu, PhD. "Dietary pattern analysis: a new direction in nutritional epidemiology", Department of Nutrition, Harvard School of Public Health, 2002, pp.1-8

[3] Martin Sarnovsky, Peter Bednar and Miroslav Smatana. "Big Data Processing and Analytics Platform Architecture for Process Industry Factories", January 2018, pp.1-2

[4] Eveline van Stijjjn, David Hesketh, Yao-Hua Tan, Bram Klievink, Sietse Overbeek, Frank Heijmann, Markus Pikart, Tim Butterly. "The Data Pipeline", January 2011, pp.4

[5] "MongoDB", Jul 2019, Retrieved from http://vuir.vu.edu.au/35038/

[6] Ajay Agarwal, Minakshi Chauhan. "Similarity Measures used in Recommender Systems: A Study", June 2017, pp.619, pp.621

[7] Dr A. Kumar. "A Fast and New Collaborative Web Recommendation System Using Fast Adaptive Association Rule Mining", 2014, pp.2-3

[8] Rakesh Agrawal, Ramakrishnan Srikant. "Fast Algorithms for Mining Association Rules", September 1994, p487-489