

# A Personalized Dietary Coaching Method Using Food Clustering Analysis

Yoori Oh<sup>†</sup> · Jieun Choi<sup>††</sup> · Yoonhee Kim<sup>†††</sup>

## ABSTRACT

In recent times, as most people develop keen interest in health management, the importance of cultivating dietary habits to prevent various chronic diseases is emphasized. Subsequently, dietary management systems using a variety of mobile and web application interfaces have emerged. However, these systems are difficult to apply in real world and also do not provide personalized information reflective of the user's situation. Hence it is necessary to develop a personalized dietary management and recommendation method that considers user's body state information, food analysis and other essential statistics. In this paper, we analyze nutrition using self-organizing map (SOM) and prepare data about nutrition using clustering. We provide a substitute food recommendation method and also give feedback about the food that user wants to eat based on personalized criteria. The experiment results show that the distance between input food and recommended food of the proposed method is short compared to the recommended food results using general methods and proved that nutritional similar food is recommended.

**Keywords :** Dietary Coaching Method, Personalized, Clustering Analysis

## 음식 군집분석을 통한 개인맞춤형 식이 코칭 기법

오 유리<sup>†</sup> · 최 지 은<sup>††</sup> · 김 윤 희<sup>†††</sup>

## 요 약

현대인의 건강관리에 대한 관심이 증가하고 다양한 만성질환을 야기하는 식습관에 대한 중요성이 강조되고 있는 상황이다. 이에 따라 여러 가지 모바일 및 웹시스템을 이용한 식단 관리 방법이 등장하고 있지만 이는 실제로 적용하기 어렵고 사용자의 상황을 반영하는 맞춤형 정보를 제공하지 않는다. 따라서 개인의 신체정보 및 상황을 반영하고 음식을 분석하여 실질적으로 사용자가 섭취 가능한 맞춤형 식단관리 및 추천 방법이 필요하다. 본 논문에서는 자기조직화지도도를 이용하여 음식을 분석하고 이를 군집화하여 음식에 대한 데이터를 준비한다. 그리고 사용자의 신체정보 및 상황을 고려한 개인 맞춤형 기준을 반영하여 섭취하고 싶은 음식에 대한 피드백 및 대체음식 추천방법을 제안한다. 또한 실험을 통하여 일반적인 방법을 이용한 추천된 음식결과와 비교하여 제안된 방법의 입력 음식과 추천 음식의 거리가 짧은 것을 통하여 영양적으로 유사한 음식이 추천됨을 증명하였다.

**키워드 :** 식이 코칭 기법, 개인맞춤형, 군집 분석

## 1. 서 론

현대인의 평균 수명이 늘어남에 따라 건강관리에 대한 관심이 증가하고 있다. 특히, 현대인의 다양한 질병을 초래하는 원인 중 하나인 식습관에 대한 중요성이 강조되고 있다. 질병관리본부의 연구에서는 당뇨, 고혈압 등 만성질환과 전반적인 식사의 질과의 연관성을 보여주고 있다[1]. 또한 이를 바탕으로 만성질환의 예방 및 관리에 있어서, 각 질환에

따른 올바른 식품선택방법 등의 식생활 관리 가이드라인이 필요하다고 한다. 이를 반영하듯 다양한 식습관 관리 방법들이 등장하고 있지만 바쁜 현대인들에게 매일 식단을 관리하고 올바르게 개선하기란 쉽지 않다.

최근 스마트폰을 이용하는 사람들이 증가함에 따라, 모바일을 이용해 쉽고 빠르게 식단을 관리할 수 있는 다양한 어플리케이션들이 등장하고 있다. 기존의 식단 관리를 위한 모바일 어플리케이션들은 공통적으로 사용자로부터 식단을 입력받고, 입력받은 식단을 바탕으로 열량, 3대 영양소의 기준치 대비 섭취 비율을 알려준다. 어플리케이션에 따라 월별 식단 평가, 건강 정보 제공 등 추가 기능도 제공되기도 하지만 대부분 다이어트 사용자를 대상으로 하다 보니 제공하는 정보가 수동적이고 큰 차이점이 없다. 사용자는 반복

<sup>†</sup> 준 회 원 : 숙명여자대학교 컴퓨터과학부 석사과정

<sup>††</sup> 준 회 원 : 숙명여자대학교 컴퓨터과학부 석사

<sup>†††</sup> 종신회원 : 숙명여자대학교 컴퓨터과학부 교수

Manuscript Received : January 5, 2016

First Revision : February 18, 2016

Accepted : February 22, 2016

\* Corresponding Author : Yoonhee Kim(yulan@sookmyung.ac.kr)

되고 사용할 때마다 유사하게 제공되는 정보들을 통하여 본인의 식습관 개선의 필요성 및 흥미를 느끼기 어렵고, 따라서 지속적인 사용에 걸림돌이 된다. 이를 개선하기 위해 본 연구에서는 사용자의 섭취기록 및 개인의 정보를 이용하여 해당 사용자에게 맞춤형 추천정보를 제공하는 기법을 소개한다. 사용자가 섭취하고 싶은 음식에 대한 섭취가능여부 응답을 제공하고 해당 음식을 섭취할 수 없는 경우, 대체음식 추천을 통해 건강한 식습관 개선을 유도하고자 한다.

본 논문의 구성은 다음과 같다. 1장의 서론에 이어 2장에서는 관련 연구들을 살펴보고, 3장에서는 제안된 방법의 분석방법 및 알고리즘에 대해 설명한다. 마지막으로 4장에서 결론을 맺는다.

## 2. 관련 연구

사용자의 식습관 코칭 관련 연구로는 [2-3]이 존재한다. [2]은 당뇨병 환자의 식습관 개선을 위한 어플리케이션 서비스 연구를 진행했다. 해당 논문에서는 설문조사를 통한 정해진 식단을 제공할 뿐, 실제 사용자가 현재 섭취한 음식이나, 섭취하고자 하는 음식을 고려하지 않는다. [3]은 식단 코칭 연구로 고혈압 환자의 비만관리를 위한 인터넷 기반 코칭 프로그램이다. 전자우편으로 사용자로부터 섭취 식단을 받고, 한국영양학회 영양분석 프로그램으로 분석한 뒤 개인별 체격지수에 맞는 섭취 칼로리, 조절 칼로리, 나트륨, 지방 섭취 조절에 대한 정보와 사용자의 부적절한 식이섭취에 대한 적절한 식이요법을 인터넷을 통해 제공한다. 그러나 [3]은 사용자가 전자우편을 매일 보내야 하는 단점과, 단순히 식이요법에 대한 지식을 높이는 데 그친다는 한계가 있다.

또한 본 논문에서 제시한 시스템과 유사한 음식 추천 연구로는 [4-6]이 존재한다. 모바일을 이용해 식단을 추천하는 연구인 DietAdvisor[4]는 스마트폰 센서를 이용하여 사용자의 데이터를 자동으로 수집하며 저장된 데이터를 바탕으로 사용자의 식성에 맞는 식단을 추천한다. 이는 사용자의 선호를 반영하지만 영양학적으로 칼로리에 대한 분석 및 정보를 제공하여 비만 이외의 질환에는 도움을 주기 어렵다. [5]는 음식의 영양소를 활용하여 자기조직화지도(Self-Organizing Map: SOM) 훈련을 통해 가중치를 구한다. 구한 음식의 가중치를 이용해 유사 특징을 가진 그룹으로 군집화하여 사용

자가 선택한 음식에 대해 유사한 음식을 추천해준다. 사용자가 선택한 음식에 대해 알고리즘을 통한 가장 비슷한 대체 음식을 추천해준다는 특징이 있지만, 사용자의 식습관이나 음식에 대한 선호도를 반영하지 않고, 당뇨병 환자에만 초점이 맞춰져 있다는 한계가 있다. [6]은 퍼지 마크업 언어를 기반으로 음식점을 추천해주는 시스템을 제안했다. [6]은 추천 시스템 온톨로지를 구성하고, 음식의 열량, 음식점까지의 거리, 음식의 가격을 퍼지화하였다. 이에 따라 열량이 낮고, 음식의 가격이 낮고, 음식점까지의 거리가 멀수록 추천 레벨이 높아진다. 열량 이외에 가격 및 음식점까지의 거리를 고려하였지만, 퍼지화하는 과정에서 범위에 대한 정확한 분석이 적용되지 않았으며, 음식점까지의 거리를 요소로 적용했다는 점이 아쉽다. 해당 논문에서는 음식점까지의 거리가 멀수록 운동을 많이 하여, 소모 열량 높다고 가정하였으나 이는 모든 경우에 적용하기 어렵다.

본 논문에서는 사용자의 섭취 이력을 바탕으로 사용자의 먹고 싶은 음식에 대해 섭취가 불가능할 경우, 먹고 싶은 음식과 유사한 대체음식을 제공하여 좀 더 나은 식습관을 유도하는 식이 코칭 방법을 제안하고자 한다.

## 3. 사용자 맞춤형 음식 추천 기법

### 3.1 데이터준비

본 연구는 식품의약품안전처의 식품영양성분데이터베이스 [7] 및 브랜드 홈페이지의 음식의 영양데이터를 기반으로 데이터베이스를 구성한다. 데이터베이스의 각 레코드는 음식 별로 번호, 해당 음식군, 주재료, 제공단위, 1회제공량(g), 열량(kcal), 탄수화물(g), 단백질(g), 지방(g), 나트륨(mg), 포화지방산(g), 칼슘(mg), 철분(mg)의 정보를 포함한다. 총 음식 개수는 887개이며, 한국에서 쉽게 접할 수 있는 음식으로 이루어져 있다.

### 1) 표준음식 데이터정보

표준음식 데이터는 22개의 그룹으로 구성되어 있으며, 유사 주재료를 사용한 음식은 같은 그룹에 속한다. 그룹은 견과류, 곡류 및 그제품 등으로 구성되어 있으며 Fig. 1은 표준음식데이터베이스의 일부를 보여주고 있다.

number	Group	New Group	Item	RECIPE	unit	serving_size	energy	carbohydrate	protein	fat	na	s_fat	ca	iron
1	곡류및그제품	밥류	쌀밥	쌀밥	공기	265	322	70.6	5.7	0.4	7.1	0	5.3	1.06
2	곡류및그제품	밥류	잡곡밥(종밥포함)	잡곡밥	공기	250	329	69.5	7.2	1.2	7.4	0.027	38.75	5.2
3	곡류및그제품	밥류	잡곡밥(종밥포함)	현미밥	공기	250	348	73.3	7.2	1.8	13.7	0	5.6	1.7
4	곡류및그제품	밥류	잡곡밥(종밥포함)	혼합잡곡밥	공기	250	346.1	76.22	7.14	0.63	6.2	0.18	34.83	3.63
5	곡류및그제품	밥류	잡곡밥(종밥포함)	종밥	공기	250	338	69.8	7.9	1.6	6.9	0	21.3	1.6
6	곡류및그제품	밥류	잡곡밥(종밥포함)	흑미밥	공기	250	348	73.3	7.2	1.8	13.7	0	5.6	1.7
7	곡류및그제품	밥류	잡곡밥(종밥포함)	보리밥	공기	260	363	79.7	7	0.4	7.6	0	8.89	0.13
8	곡류및그제품	밥류	잡곡밥(종밥포함)	찰밥	공기	250	359.65	78.82	6.57	0.39	6.47	0.16	15.1	3.83
9	곡류및그제품	밥류	비빔밥,볶음밥	비빔밥	공기	750	637	98.3	22.4	15.2	714.9	5.38	182.33	4.05
10	곡류및그제품	밥류	김밥	김밥	줄	200	317.68	57.57	7.26	6.48	833.29	2.75	59.36	1.02
11	곡류및그제품	밥류	비빔밥,볶음밥	볶음밥	공기	200	224	38.1	6.6	4.1	258.2	6.39	35.06	0.94
12	곡류및그제품	밥류	비빔밥,볶음밥	김치볶음밥	공기	300	312	56.7	7.6	4.8	127.5	4.26	89.13	1.38
13	곡류및그제품	회,조밥류	회,조밥류	생선조밥	접시	300	461.97	76.43	25.42	6.06	968.94	1.51	45.36	0.57

Fig. 1. Standard Food Database

2) 만성질환별 유의영양소 권장량

현대인의 주의해야 할 4가지 만성질환인 고혈압, 당뇨, 고지혈증, 비만을 선정하여 각 만성질환의 유의 영양소에 대한 권장량 기준을 선정한다. 고혈압은 나트륨, 고지혈증은 포화지방산, 당뇨는 탄수화물, 비만은 열량을 유의 영양소로 판단한다. 또한 권장량은 [8]에서 제안하는 사용자의 성별, 나이, 키, 몸무게, 활동계수로 계산되는 에너지필요추정량 Equation (1)을 기반으로 정해진다.

$$E(\text{에너지추정량 : Male}) = 622 - 9.53 * \text{Age}(\text{year}) + PA[15.91 * \text{Weight}(\text{kg}) + 536.6 * \text{Height}(\text{m})]$$

$$PA = 1.1(\text{저활동적}), 1.48(\text{매우 활동적})$$

$$E(\text{에너지추정량 : Female}) = 354 - 6.91 * \text{Age}(\text{year}) + PA[9.36 * \text{Weight}(\text{kg}) + 726 * \text{Height}(\text{m})]$$

$$PA = 1.2(\text{저활동적}), 1.45(\text{매우 활동적})$$

에너지추정량은 비만의 유의 영양소인 열량의 권장량 기준이 된다. 도출한 에너지추정량을 기반으로 다른 만성질환의 유의영양소 권장량을 계산할 수 있다. 각 만성질환별, 유의영양소의 권장량 계산방법은 Table 1과 같다[8]. 고혈압의 유의영양소인 나트륨의 경우, 한국인은 나트륨을 과잉섭취 경우가 많기 때문에 권장량(2000mg)이 아닌 식품의약품안전처에서 제시한 목표섭취량인 3700mg을 기준으로 선정한다.

Table 1. Recommended Amount for each Chronic Disease and Caution Nutrients

만성질환	유의 영양소	권장량
고혈압	나트륨	3700(mg)
당뇨	탄수화물	$E * 0.7/4(\text{g})$
고지혈증	포화지방산	$E * 0.07/9(\text{g})$
비만	열량	$E(\text{kcal})$

3.2 음식 군집 분석

음식은 6가지 주요 영양소(열량, 탄수화물, 단백질, 지방, 포화지방산, 나트륨)으로 군집분석을 수행한다. 해당 영양소는 부족 또는 과잉 섭취 시, 질환을 일으킬 수 있는 위험이 있으므로 선택하였고, 특히, 만성질환에 따른 위험을 판단하기 위해 4가지 영양소에 대해서는 사용자 맞춤형 권장량을 계산하여 부족 및 과잉 섭취여부를 판정한다. 또한 모든 음식 데이터는 훈련을 진행하기 전에 표준화된다.

군집분석은 두 가지 단계로 이루어진다. 처음에는 자기조직화 지도(Self-Organizing Map: SOM)을 만들어 훈련을 진행하고 자기조직화 지도에 k-평균 군집분석을 이용하여 군집을 형성한다. 마지막 결과는 각 군집의 음식들은 영양적으로 비슷한 구성을 가진 음식으로 군집화 된다.

1) 자기조직화 지도 훈련

본 연구에서는, R의 "kohonen" 패키지[9]를 사용하여 구현하였다. 자기조직화 지도는 보통 U-Matrix[10]를 이용하여 결과를 표현한다. Fig. 2에서 자기조직화지도의 노드는

각 노드를 구성하는 가중치 값을 가지고 있다. 또한 좌측 축은 노드의 각 색깔 별로 해당하는 값을 나타낸다.

Fig. 3은 가중치 벡터를 구성하는 변수인 열량, 탄수화물, 단백질, 지방, 포화지방산, 나트륨의 정규화를 적용하기 전 분포를 나타낸다. 노드의 색깔을 통해 각 영양소별 구성이 다를 수 있다.

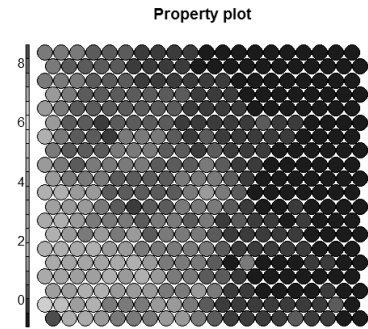


Fig. 2. U-matrix Visualization of the SOM

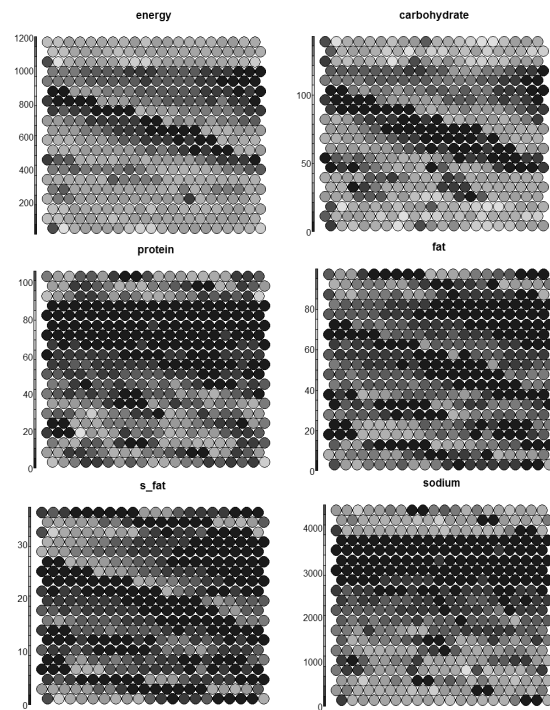


Fig. 3. Visualization of the SOM of Nutritive Values for Food Dataset

2) k-평균 군집분석

자기조직화 지도에서 도출된 각 노드는 영양소별 특징 벡터를 속성으로 가진다. 벡터는 열량, 탄수화물, 단백질, 지방, 포화지방산, 나트륨의 정규화된 값으로 구성되어 있다. 예를 들어, '굴'의 영양소 벡터는 [-1.07, -0.74, -0.71, -0.94, -0.80, -0.87]이며, '떡갈비'의 영양소 벡터는 [2.66, -0.30, 3.20, 4.64, 0.42, 4.53]이다. 이를 k-평균 군집분석을 이용하여 6개의 군집으로 군집화하였다. k 값을 정하기 위하여, Elbow 방법

[11]을 사용하였다. 군집을 2개부터 15개의 그룹으로 형성하였을 때, 군집에 포함된 노드와 중심까지 거리의 제곱합을 계산하여 급격히 변하는 점인  $k=6$ 을 선택하였다. Fig. 4는 군집결과를 보여준다.

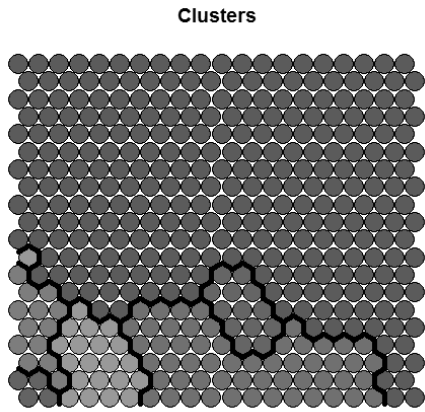


Fig. 4. Visualization of the Clustered Map

### 3.3 질문 알고리즘

```

Algorithm 1. Q&A
Input: Nutrition List of Food  $F[]$ , Name of Food  $F_{name}$ 
Output: Name of Suggested Food  $F_{suggestion}$ , Amount of Suggested Food  $F_{amount}$ ,
Name of Overtaken Nutrition  $Over_{nutrition}$ , Amount of overtaken Nutrition  $Over_{amount}$ 
1:  $Recom[] \leftarrow$  CalculateRDA(sex, height, weight, PA);
2:  $Intake[] \leftarrow$  SumOfIntake(CurrentTime);
3: switch(meal){
4:   Case breakfast || lunch :
5:      $Criterion \leftarrow Recom$ ;
6:   Case dinner :
7:      $Criterion \leftarrow Recom - F$ ;
8: }
9: for each nutrition in  $Recom$  do
10:  If  $Intake[nutrition] < Criterion[nutrition]$  then
11:     $F_{suggestion} \leftarrow F_{name}; F_{amount} \leftarrow \alpha$ ;
12:  else if  $Intake[nutrition] < Criterion[nutrition] * \gamma$  then
13:     $F_{suggestion} \leftarrow F_{name}; F_{amount} \leftarrow \beta$ ;
14:  else
15:     $over_{nutrition} \leftarrow$  nutrition;
16:     $over_{amount} \leftarrow Intake[nutrition] - Recom[nutrition]$ ;
17:     $F_{suggestion} \leftarrow$  Min(Distance(F)) where  $F_{suggestion}[nutrition] < F_{nutrition}$ ;
18:     $F_{amount} \leftarrow \alpha$ ;
19:  end if
20: end for
    
```

<Algorithm 1. Q&A >

알고리즘 1은 사용자의 선호 음식의 적합성 및 대체 음식을 제안한다. 따라서 입력된 음식을 분석한 결과로, 섭취가능 여부와 섭취 불가능시, 과잉 섭취한 영양소에 대한 정보 및 대체음식을 출력한다.

먼저, 사용자의 신체 및 개인정보를 이용하여 3.1의 2) 만성질환별 유의영양소 권장량에서 언급된 수식으로 4가지 영

양소(열량, 탄수화물, 포화지방산, 나트륨)의 권장섭취량을 계산한다(line 1). 또한 현재 시간에 따라 특정 식사의 영양소별 총 섭취량을 계산한다. 아침 또는 점심일 경우, 어제 하루의 식사를 반영하고, 저녁일 경우, 당일 아침과 점심의 식사를 반영한다(line 2). 마찬가지로, 아침 또는 점심일 경우, 판단기준은 권장량을 그대로 적용하고, 저녁일 경우, 권장량에서 사용자가 섭취하고 싶은 입력 음식의 영양소의 차를 기준으로 적용한다(line 4-7). 앞서 도출한 정보들과 3.2의 군집을 이용하여 결과를 출력한다. 권장영양소 각각에 대해 판단기준에 대한 해당 범위에 따라 다른 결과를 나타낸다. 특정영양소의 섭취량이 판단기준보다 작을 경우, 해당 음식을 제안하고, 그 양을  $\alpha$ 로 제안한다(line 11). 영양소의 섭취량이 판단기준을 초과하였지만 일정 비율  $\gamma$ 를 넘지 않는 범위에 있다면, 해당음식을 제안하지만, 그 양을 줄인  $\beta$  ( $\alpha > \beta$ )로 제안한다(line 13). 영양소 섭취량이 일정 비율  $\gamma$ 을 넘는 과잉 범위에 있다면, 대체 음식을 제안한다. 우선, 과잉 섭취한 영양소 및 과잉 섭취량을 출력한다. 또한 3.2의 군집의 노드 중, 섭취하고 싶은 음식의 노드에서 거리가 가장 작은 음식을 선택한다. 이때, 선택된 음식의 영양소는 먹고 싶은 음식의 영양소보다 작은 값을 가져야한다(line 15-18). 따라서 사용자가 먹고 싶은 음식에 대해 개인 정보를 활용하여 섭취가능여부를 제공하고, 맞춤형 대체음식을 제공한다.

### 3.4 실험

3.3의 알고리즘을 적용하여 사용자가 먹고 싶은 음식을 질문하고 해당 음식이 섭취 불가능한 경우 대체음식을 제공한다. 이때, 3가지 실험을 진행하였다. 대체음식을 제공 시, 음식군집분석 후, 군집 내에서 최소의 거리를 갖는 음식, 군집에 관계없이 최소의 거리를 갖는 음식, 일반적인 음식군내에서의 랜덤 기법으로 추출한 음식으로 각각 방법을 사용하여 대체음식으로 추천하고 그 결과를 비교하였다. 측정하는 거리는 두 음식 벡터의 거리를 구하므로 유클리드 거리를 이용하였다. 랜덤 기법이란, 알고리즘의 전체적인 흐름은 같으나, 대체음식을 제공할 때 같은 음식군의 다른 조리법을 가지며 과잉섭취 영양소가 적게 함유된 음식을 추천하는 것을 의미한다. 만약 그러한 음식이 존재하지 않으면, 섭취하고 싶은 음식이 속한 음식군에서 무작위로 선택하여 추천한다.

사용자가 포화지방산을 과잉 섭취한 경우, 섭취하고 싶은 음식을 질문했다고 가정하자. 다음 Table 2는 입력한 음식에 따라 총 3가지 방법, 음식군집분석 후, 군집 내에서 최소 거리를 갖는 음식, 군집에 관계없이 최소의 거리를 갖는 음식, 일반적인 음식군내에서의 랜덤 기법으로 추천한 음식을 결과 음식으로 나타내고, 입력음식과 결과음식의 정규화된 거리를 나타낸다. 해당 그룹에서 추천할 음식이 없을 경우, 음식 추천결과가 없는 경우도 있다. 결과를 비교하였을 때, 군집분석을 적용했을 경우, 추천결과음식은 입력음식과 영양적으로 유사한 음식이 추천되었음을 알 수 있다. '오렌지 주스'를 입력했을 때, 군집분석을 적용한 경우, '귤'을 추천해

Table 2. Recommended Food Results for Experiments

	군집분석-군집 내에서 추천		군집분석-군집 구분 없이 추천		일반적인 음식군-랜덤 기법	
	결과음식	거리	결과음식	거리	결과음식	거리
입력음식	결과음식	거리	결과음식	거리	결과음식	거리
오렌지주스	귤	0.0216	귤	0.0216	아리조나아이스티	0.5491
찰밥	보리밥	0.0652	보리밥	0.0652	찜감자	2.4246
절편	증편	0.0848	증편	0.0848	찜옥수수	0.7067
떡갈비	오리훈제구이	1.914	치즈돈가스(치즈돈까스)	1.8512	두부김치	6.8409
베이컨토마토디럭스 (맥도날드)	빅맥(맥도날드)	0.3855	빅맥(맥도날드)	0.3855	불고기버거(삼강)	2.3228
치킨까스 (치킨커틀렛)	칸풍기	0.4514	씨제이프레시안백설잡채 군만두(씨제이제일제당)	0.3024	맥너겟(맥도날드)	3.9237
육개장라면	새우탕큰사발면	0.1652	새우탕큰사발면	0.1652	꿀떡	4.285
칼국수	후루룩국수	0.5854	후루룩국수	0.5854	메밀소바	1.1148
새우튀김롤	송이덮밥	0.7142	송이덮밥	0.7142	두부김치	3.2868
한방오리백숙	X		케이준치킨버거(삼강)	1.1544	돼지고기메추리알장조림	5.7836
참치김밥	샐러드김밥	0.4167	샐러드김밥	0.4167	참치마요네즈삼각김밥	2.3477
사천짜파게티큰사발	찰비빔면	0.3205	찰비빔면	0.3205	카레라이스	2.5044
쌀국수설렁탕	메밀소바	0.2308	메밀소바	0.2308	참쌀떡	3.0522
평균 거리		0.4462		0.4844		3.0109

준다. 반면, 일반적인 음식군의 랜덤 기법 적용 결과음식은 ‘아리조나아이스티’가 도출되었다. 또한 ‘떡갈비’를 입력하면, 분석한 군집내 최소거리를 갖는 음식으로 ‘오리훈제구이’가 도출되고, 군집을 구분하지 않으면 영양적으로 유사한 ‘치즈돈가스(치즈돈까스)’가 도출되었다. 한편, 일반적인 음식군의 랜덤기법을 이용한 결과로는 ‘두부김치’가 도출되었다. 이는 군집분석을 적용하였을 때, 일반적인 음식군을 이용하여 추천했을 때보다 더 영양적으로 유사한 음식이 추천되었음을 알 수 있다. 한편, ‘한방오리백숙’을 입력했을 경우, 분석한 군집 내 최소거리를 갖는 음식이 존재하지 않아 결과음식이 없었지만, 군집 구분없이 추천했을 때, ‘케이준치킨버거(삼강)’이 추천되었다. 일반적인 음식군에서 추천을 했을 때는 ‘돼지고기메추리알장조림’이 결과로 나왔지만 이는 입력음식과의 거리가 ‘케이준치킨버거(삼강)’에 비해 약 5배 거리가 먼 것으로 나타났다. 이는 군집내에서 최소거리의 음식이 없을 시, 군집을 구분 없이 추천한다면 일반적인 음식군을 적용할 때보다 입력음식과 연관성이 더 큰 음식을 추천할 수 있고, 다양한 입력음식에 대해 유사한 추천음식을 제공할 수 있음을 증명하였다. 정규화된 음식 집합에서 입력음식과 결과음식간의 거리를 계산했을 때, 각각의 경우 평균 거리는 0.4462, 0.4844, 3.0109를 가지며, 군집분석을 적용한 추천음식결과에서 음식간의 평균 거리가 짧음을 알 수 있다. 이는 영양적으로 유사한 음식이 추천되었음을 의미한다. 또한 군집분석 후, 군집내에서 추천하는 경우보다 군집을 넘어서는 것을 허용했을 때 결과음식의 평균거리가 더 크게 나타나는 것은 일부 음식은 더 거리가 더 짧은 음식을 추천 하였지만, 군집내에서는 추천이 되지 않았던 음식이 추천되는 경우가 존재하기 때문이다. 따라서, 각 경우의 입력음식과 결과음식의 평균거리를 계산하였을 때, 군집분석을 적

용했을 때, 일반적인 음식군을 이용했을 때보다 약 7배정도 거리가 더 먼 음식을 추천하였고 이를 통하여 사용자가 입력한 섭취하고 싶은 음식에 대해 유사한 음식을 추천해주었음을 증명하였다.

#### 4. 결 론

본 논문에서는 음식의 영양소 데이터를 군집 분석하여 사용자가 섭취하고 싶은 음식에 대한 대체음식 추천방법을 제안한다. 음식데이터를 자기조직화 지도를 이용하여 영양적으로 분석하고 k-평균 클러스터링을 이용하여 군집화한다. 또한 사용자의 섭취한 식단인력을 분석하고 사용자의 신체 정보 및 개인정보를 이용하여 개인의 영양소별 권장량을 계산한다. 분석된 정보를 이용하여 사용자가 먹고 싶은 음식을 질문하였을 때, 해당 음식의 섭취가능여부 및 양을 제공하고 섭취 불가능한 경우, 대체음식을 제안한다. 대체음식은 사전에 분석한 음식 클러스터를 활용하여 먹고 싶은 음식과 유사한 음식을 제안한다. 또한 세 가지 경우의 실험을 통하여 군집분석을 적용했을 때, 영양적으로 유사한 음식을 추천해주었음을 증명하였다. 첫 번째는 군집분석결과의 군집내에서 최소 거리를 갖는 음식을 추천하고, 두 번째는 군집 분석결과의 군집을 넘어서 추천을 허용하고, 마지막 실험은 일반적인 음식군에서 랜덤 기법을 이용하여 추천했을 때 결과를 나타낸다. 이를 통하여, 군집분석을 적용한 결과와 일반적인 음식군을 이용한 결과를 비교하여 군집분석을 적용한 결과가 입력 음식과 더 영양적으로 유사한 음식을 추천 해주었음을 증명하였다. 제안된 방법을 통해 사용자는 개인 맞춤형 대체음식 추천 정보를 얻고, 보다 현실적이고 실제 식습관에 적용 가능한 결과를 얻을 수 있다. 따라서 사용자

는 추천된 음식을 섭취함으로써 건강한 식습관을 가지고, 만성질환을 예방하는 효과를 얻을 수 있다. 향후에는 영양소뿐만 아니라 음식의 가격, 현재 날씨, 사용자의 위치를 기반으로 다양한 정보를 반영하여 음식추천 알고리즘을 확장할 계획이다.

### 감사의 글

본 논문에서의 음식 데이터베이스 설계 및 기능별 영양학적 판단에 숙명여자대학교 식품영양학과 이정은 교수님과 이지은 대학원생의 도움을 받음.

### References

[1] J. E. Yu and Y. J. Kim, "Healthy Eating Index (HEI) Indicates that High Diet Quality is Associated with Low Prevalence of Hypertension and Type 2 Diabetes in Koreans: the Korean Genome and Epidemiology Study (KoGES)," *Public Health Weekly Report*, Vol.8, No.3, pp.51-58, 2015. (in Korean).

[2] Jongsu Lee, "A Study on Application Service for Dietary Habit Improvement of Patients with Chronic Disease - Focused on Patients with Hypertension and Diabetes -," *Journal of Korea Society of Design Forum*, Vol.48, pp.71-82, 2015. (in Korean).

[3] Hae-Ok Jeon and Ok-Soo Kim, "The effects of an internet based coaching program for obesity management in hypertensive patients," *Korean Journal of Adult Nursing*, Vol.23, No.2, pp.146-159, 2011.

[4] B. K. Lim, et al., "DietAdvisor: A Personalized eHealth Agent in Mobile Computing Environment," *Conference on KIISE*, Vol.38, No.2(D), pp.115-118, 2011.

[5] Phanich, Maiyaporn, Phathrajarin Pholkul, and Suphakanth Phimoltares, "Food recommendation system using clustering analysis for diabetic patients," *Information Science and Applications (ICISA), 2010 International Conference on. IEEE*, 2010.

[6] Woan-Tyng Lin, et al., "FML-Based Recommender System for Restaurants," *Technologies and Applications of Artificial Intelligence (TAAI), 2013 Conference on. IEEE*, 2013.

[7] Food Nutrients Database [Internet], <http://www.foodnara.go.kr/kisna/index.do>.

[8] The Korean Nutrition Society, Dietary Reference Intakes for Koreans, revised Version, The Korean Nutrition Society, Korea, 2010.

[9] Ron Wehrens and MC Buydens Lutgarde, "Self-and super-

organizing maps in R: the Kohonen package," *Journal of Statistical Software*, Vol.21, No.5, pp.1-19, 2007.

[10] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 1, No.4, pp.224-227, 1979.

[11] Elbow Method [Internet], [https://en.wikipedia.org/wiki/Determining\\_the\\_number\\_of\\_clusters\\_in\\_a\\_data\\_set](https://en.wikipedia.org/wiki/Determining_the_number_of_clusters_in_a_data_set).



### 오 유 리

e-mail : yoori0203@sookmyung.ac.kr  
 2015년 숙명여자대학교 수학과(학사)  
 2015년~현재 숙명여자대학교 컴퓨터  
 과학부 석사과정  
 관심분야: 빅데이터 분석, 지능형시스템



### 최 지 은

e-mail : jechoi1205@sookmyung.ac.kr  
 2014년 숙명여자대학교 컴퓨터과학부(학사)  
 2016년 숙명여자대학교 컴퓨터과학부 석사  
 관심분야: 분산시스템, 클라우드 컴퓨팅



### 김 윤 희

e-mail : yulan@sookmyung.ac.kr  
 1991년 숙명여자대학교 전산학과(학사)  
 1996년 Syracuse University 전산학과  
 (석사)  
 2000년 Syracuse University 전산학과  
 (박사)  
 1991년~1994년 한국전자통신연구소(ETRI) 연구원  
 2000년~2001년 Rochester Institute of Technology 컴퓨터공학과  
 조교수  
 2001년~2004년 숙명여자대학교 컴퓨터과학과 조교수  
 2004년~2009년 숙명여자대학교 컴퓨터과학과 부교수  
 2009년~현재 숙명여자대학교 컴퓨터과학부 교수  
 관심분야: 그리드/클라우드 컴퓨팅 환경, 워크로플로우 제어,  
 그리드/클라우드 관리