

# 감성 분석을 이용한 다이어트 보조 식품에 대한 온라인 평판 분석

이소희, 이진영, 김현희  
동덕여자대학교 정보통계학과  
e-mail : [dlthgml1230@naver.com](mailto:dlthgml1230@naver.com)  
[solia0394@naver.com](mailto:solia0394@naver.com)  
[heekim@dongduk.ac.kr](mailto:heekim@dongduk.ac.kr)

## Online Reputation Analysis of Dietary Supplements based on Sentiment Analysis

So-Hee Lee, Jin-Yeong Lee, Hyon Hee Kim  
Dept. of Statistics and Information Science, Dongduk Women's University

### 요 약

본 연구에서는 체중 감량을 위해 무분별한 다이어트 식품의 남용을 막고, 다이어트 보조 식품에 대한 정보를 제공하기 위해서 감성 분석을 활용하여 다이어트 보조 식품에 대한 온라인 후기를 분석하였다. 먼저, 다이어트 보조 식품을 그 특성에 따라 네 가지 종류로 분류하고 각 카테고리 별로 긍정 및 부정 점수를 계산하였다. 이를 위해 체중 감량에 대한 감성 사전을 다이어트 식품에 대한 후기를 텍스트 마이닝하여 구축하였다. 특히 부작용이 있는 식품에 대한 부정 점수에 가중치를 두기 위해서 WHO-ART 에서 정의한 부작용 용어에는 가중치를 두어 처리하였다. 분석 결과 단백질 보충 식품군이 긍정 점수가 가장 높게 나타났고, 이는 다이어트를 위한 목적 이외에도 운동을 전문적으로 하는 사람들에게 오랜기간 사용되어 왔기 때문인 것으로 해석된다. 또한 식욕 억제제 식품군이 긍정점수는 가장 낮고 부정 점수는 가장 높게 나타났는데, 이는 식욕억제제의 주성분인 펜타민에 의한 가능성이 클 것이라고 예측된다.

### 1. 서론

최근 들어 한국 사회에서 건강상의 이유뿐만 아니라 외형적인 변화를 주기 위한 방법으로 다이어트를 많이 시도한다. 그에 따라 다이어트 보조 식품과 관련된 수요가 증가하고, 자연스레 시장이 급격하게 커지며 수많은 제품들이 출시되었다. 그러나 수많은 제품들이 출시되며 과장 광고 및 허위 광고를 하고, 이에 속아 다이어트 보조 식품을 섭취했다가 부작용을 겪은 사례도 증가하고 있다. 따라서 이러한 다이어트 보조 식품들의 후기를 분석하여 무분별한 남용을 방지하고자 했다. [1,2]

본 연구는 포털 사이트를 크롤링(crawling)하여 다이어트 보조 식품에 관한 후기 및 댓글들을 수집한 뒤, R 프로그램에서 감성 분석을 실시하여 다이어트 보조 식품의 그룹별 온라인 평판 분석을 실시하였다. SNS 댓글과 후기를 참고하여 다이어트를 주제로 한 감성 사전을 구축하였으며, 특히 World Health Organization Adverse Reaction Terminology (이하 “WHOART”라 함)에 정의된 부작용을 나타내는 표준 단어들에 가중치를 부과하여 부정 점수를 산정하였다. [3]

다이어트 보조 식품은 섭취 후 흡수 형태에 따라 4 가지 종류로 분류하였다. (A) 포만감을 느끼게 하는 식품, (B) 식욕을 억제하도록 돕는 식품, (C) 단백질을 보충해 주는 식품, (D) 칼로리 흡수를 저지하는 식품으로 분류했으며, 각 식품의 주성분은 (A) 알긴산, (B) 펜타민, (C) 프로틴, (D) 가르니시아 [4,5]였다. 실험을 위해서 포털 사이트의 사용자 블로그를 수집하였으며 구축된 감성 사전을 기반으로 각 카테고리 별로 긍정 및 부정 점수를 산정하였다. 실험 결과 가장 높은 긍정 점수를 받은 카테고리는 (C) 단백질을 보충해주는 식품, 가장 낮은 긍정 점수를 받은 카테고리는 (B) 식욕을 억제하도록 돕는 식품이었다. 또한 가장 높은 부정 점수를 받은 카테고리는 (B) 식욕을 억제하는 식품이고 가장 낮은 부정 점수를 받은 카테고리는 (D) 칼로리 흡수를 저지하는 식품이었다

본 논문은 다음과 같이 구성된다. 제 2 장에서는 데이터 수집 및 탐색에 대하여 설명하고 제 3 장에서 감성 사전 구축 기법에 대하여 기술한다. 제 4 장에서 분석 결과를 제시하고, 마지막으로 제 5 장에서 결론을 맺는다.

## 2. 데이터 수집 및 탐색

본 연구는 텍스트 마이닝(Text Mining)을 이용하여 포털 블로그에서 각 다이어트 보조 식품의 장점과 단점 및 부작용을 검색 후, 100 여 개의 사용자들의 후기를 스크랩하여 수집하였다. 수집 기간은 2018 년 2 월 5 일부터 2018 년 3 월 11 일까지 약 한달 간의 자료를 분석하였다.

분류된 4 가지 카테고리 별 후기를 크롤링(crawling)하여 분석할 문장의 군집을 살펴보았다. 그 이후 세종사전(한글 데이터를 처리하기 위해 만들어진 사전, 이 사전에 등록되지 않은 한글은 처리할 수 없다.)을 기반으로 문장에서 명사만 추출하는 함수를 사용하여 명사를 추출하였다. 그러나 빈칸이나 조사 등의 불필요한 단어가 상위 빈도수를 많이 차지하고 있었기 때문에 데이터 전처리를 통해 데이터를 정제하고, 상위 빈도 단어로 워드 클라우드를 생성하고자 하였다.

데이터 전처리는 먼저 자음과 모음으로만 이루어진 단어, 의성어, 이모티콘, 특수문자, 영어로 이루어진 단어를 삭제하였다. 그 이후 특정 단어를 빈칸으로 만들어 주는 함수를 활용하여 필요하지 않은 단어, 무분별한 동사들을 원형으로 기본 동사의 형태로 통일 시켰다. 이 두 단계를 끝내고 워드 클라우드를 그려봤지만 한 글자 단어가 많이 추출되어 워드 클라우드의 대다수를 차지하고 있었다. 그래서 단어의 길이가 최소 2 글자 이상인 단어를 추출하였다. 그 이후 최소 3 번은 나온 상위 50 개의 워드 클라우드를 생성하여 내림차순으로 정리하였다. [6]



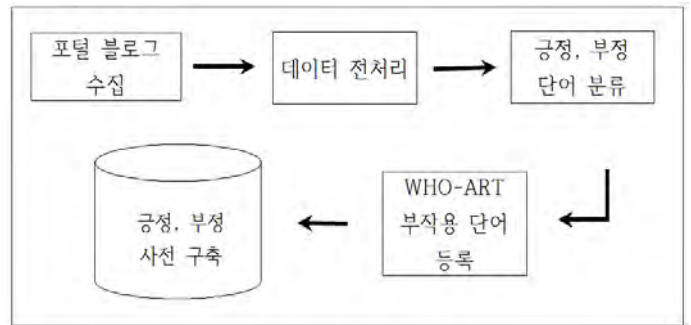
(Figure1) 각 카테고리 별 워드 클라우드

(Figure 1)은 각 카테고리 별 중요 단어를 추출한 다음 상위 50 개 단어를 워드 클라우드로 시각화 한 것이다. (A) 포만감을 느끼게 하는 식품에서는 포만감, 식욕, 변비, 가르니시아 등의 단어가 상위 50 개의 단어로 추출되었다. (B) 식욕을 억제하도록 돕는 식품은 식욕, 지방, 비타민, 구토, 당뇨병의 단어를 볼 수 있

었다. (C) 단백질을 보충해 주는 식품은 건강, 운동, 증가, 탈모 등의 단어가 추출되었다. (D) 칼로리 흡수를 저지하는 식품에서는 차단, 흡수, 마그네슘, 강력, 주의의 단어가 추출되었다.

## 3. 사전 구축

감성 사전은 비정형 데이터를 분석하여 긍정, 부정으로 나누어 활용한다. 다이어트 보조 식품에 관련한 상품평을 분석하기 위해 다이어트 보조 식품 사용자들의 어휘를 고려하여 감성 사전을 구축하였다. [7]



(Figure 2) 감성 사전 구축 process

(Figure 2)는 감성 사전 구축 과정이다. 감성 사전 구축을 위해서는 5 가지 단계로 구성된다. 가장 먼저 포털 블로그 자료를 수집하여 데이터 전처리를 통해 단어를 정제하였다. 다음 단계로 블로그나 SNS 에 게시된 다이어트와 관련된 후기를 참고하여 자주 등장하는 단어가 긍정적으로 사용되는지, 부정적으로 사용되는지 직접 판단 후 사전을 구축하였다.

<표 1> 감성 사전의 일부 예시

분류	단어
긍정	날씬, 가느다란, 추천, 변비탈출, 간편, 식욕부진, 성공, 효율적, 줄어들다, 감소하다, 빠지다, 억제, 도움, 원활, 성공, 극대화, 만족, 극복, 체중감량
부정	거짓, 증가, 별다른, 제자리, 돈낭비, 과대광고, 실망, 없다, 체중증가, 일시적, 요요, 가격부담, 부작용, 무분별, 손상, 피해, 포기, 주의, 과도, 왕성, 내성, 망했다
WHO-ART	손떨림, 두통, 탈모, 불면증, 우울증, 신경과민, 설사, 탈수, 현기증, 여드름, 소화장애, 무기력, 간손상, 구토, 생리 불순, 두드러기, 메스꺼움, 어지러움,

<표 1>은 감성 사전의 일부 예시이다. 예를 들어, SNS 게시물에 달린 ‘제품에 효과가 없어요’라는 댓글에서 ‘없다’라는 단어는 부정적인 의미로 사용되었기에 부정 사전에 등록하고, ‘0.9키로 뺐네요’ 라는 댓글에서 ‘빠지다, 빠다’의 단어는 긍정적으로 사용되었기 때문에 긍정 사전에 등록하였다. 또한 부정 점수에서 후기나 검색 결과에 부작용 사례가 등장할 때, WHO-

ART 에 등록되어 있는지 유무를 확인 후 등록되어 있다면 가중치를 주어 계산하였다.

다이어트 보조 식품의 후기를 분석하는데 사용된 감성 사전의 가장 큰 특징은 일반적으로 감성분석에서 사용되는 부정의 단어가 본 사전에서는 긍정으로 쓰이고, 긍정의 단어가 부정으로 사용된다는 것이다. <표 1>을 보면 ‘줄어들다, 식욕부진, 감소하다, 빠지다’ 등의 단어는 일반적으로 부정적인 의미를 가지지만 본 연구에서는 긍정적인 의미로 사용되었다. 반대로 부정의 예시로는 ‘증가, 왕성, 늘었다’ 등의 단어가 부정적인 의미로 사용되었다.

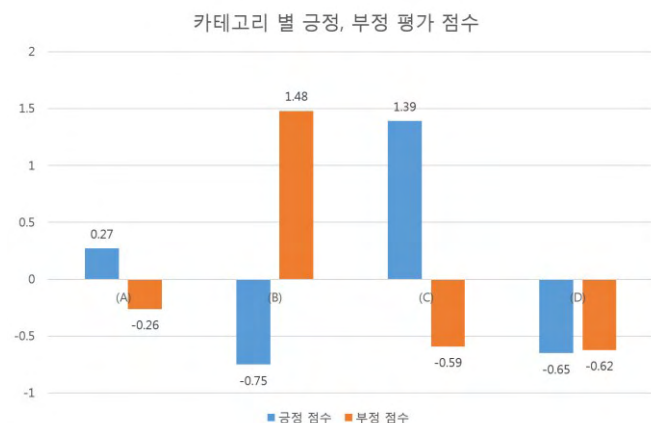
#### 4. 분석 결과

<표 2> 긍정 수치, 부정수치

분류	A	B	C	D	평균	표준 편차
긍정	12.51	8.18	17.26	8.60	11.367	4.23
부정	9.23	19.38	7.30	7.13	10.76	5.83

<표 2>는 긍정사전과 부정사전을 활용하여 각 카테고리 별로 긍정의 수치를 모두 더한 뒤 전체 데이터의 수로 나누어 긍정평가의 평균을 구하고, 같은 방식으로 부정평가의 평균을 구하였다. 그 뒤 각 카테고리의 표준편차를 구하였다.

긍정, 부정 점수의 기준치가 다르기 때문에 이후 긍정, 부정 점수를 각각 z-score 정규화하여 비교하고자 하였다. 각 카테고리 별로 긍정 점수와 부정 점수에 대한 평균과 표준 편차를 구한 뒤, 네 가지 카테고리의 원 점수들에서 각각의 평균을 뺀 이후 다시 표준편차로 나누었다. 이러한 z-score 정규화를 사용하게 되면 기준치가 같아지므로 긍정과 부정 점수의 비교가 가능하게 된다.



(Figure 3) 카테고리 별 긍정, 부정 평가 점수

(Figure 3)은 카테고리 별 긍정, 부정 평가 점수를 나타낸 그래프이다. z-score 정규화를 사용하면 점수는 (-3, 3) 사이의 점수를 나타낸다. z-score 정규화를 사용한 감성 분석 이후 평가 점수를 살펴봤을 때, 긍정

점수는 (C) 단백질을 보충해주는 식품, (A) 포만감을 느끼게 하는 식품, (D) 칼로리를 저지하는 식품, (B) 식욕을 억제하는 식품 순으로 높다.

부정 점수는 (B) 식욕을 억제하는 식품, (A) 포만감을 느끼게 하는 식품, (C) 단백질을 보충해주는 식품, (D) 칼로리를 저지하는 식품 순으로 높다.

#### 5. 결론

(C) 단백질을 보충해주는 식품의 긍정 점수가 가장 높은 이유는 다이어트를 위한 목적 이외에도 사용되어 왔기 때문이다. 운동을 전문적으로 하거나 대회에 나가는 사람들에게 근육의 생성에 도움을 주고, 그 외에 체내 에너지를 활성화하거나 피로 회복 등 다이어트 보조 식품이 각광받기 시작한 최근보다 훨씬 전에도 사용되어 왔기 때문이라고 판단된다.

부정 점수가 가장 높은 (B) 식욕을 억제하는 식품은 긍정 점수도 가장 낮았다. 그 이유는 식욕억제제의 주성분인 펜타민에 의한 가능성이 클 것이라 생각된다. 펜타민은 체질량지수가 매우 높거나 고혈압, 당뇨 등의 위험 요인이 있는 비만 환자의 치료에 사용되는 식욕억제제이다. 뇌에서 배고픔을 덜 느끼게 하거나 포만감을 증가시키는 신경전달물질의 작용을 증가시키는데, 오용이나 남용 시에 심각한 부작용이 생길 수 있기 때문에 향정신성의약품으로 분류되어 있다. [8] 이러한 펜타민 성분 때문에 부작용이 다른 세 가지 식품에 비해 많이 언급되기 때문에 부정적인 점수가 매우 높다고 본다. 따라서 식욕 억제 식품을 섭취하여 다이어트를 하고자 하는 소비자들은 보다 특별한 주의를 기울여야 할 것이다.

#### 참고문헌

- [1] 편집부, “다이어트식품 관련 소비자피해 실태조사”, 한국소비자원 조사보고서(2006): 60-79
- [2] 백병성, “인터넷을 통한 다이어트 관련 광고 실태조사 결과보고”, 한국소비자원 조사보고서(2009): 160-182
- [3] 임경화 외 7 인, “의약품 부작용에 관한 국제 분류체계인 WHO-ART 와 MedRA 의 비교 분석”, 임상약학회지 (2007), 17(1): 46-52.
- [4] 구글, “다이어트보조제분류”, 2018-03-13, <http://news.joins.com/article/18123492>
- [5] 네이버, “다이어트보조제분류”, 2018-04-04, <https://blog.naver.com/cristy82/220630832181>
- [6] 정찬훈 외 3 인, “리뷰 데이터의 신뢰도를 반영한 한국어 감성사전”, 한국정보과학회 한국소프트웨어종합학술대회 논문집(2017): 1-2.
- [7] 장경애 외 2 인, “인터넷 감성기호를 인용한 긍정/부정 말뭉치 구축 및 강진분류 자동화”, 한국정보과학회 정보과학회논문지(2015), 42(4): 513-515
- [8] 네이버, “펜타민”, 2018-04-04, <http://terms.naver.com/entry.nhn?docId=5138152&cid=59913&categoryId=59913>