

# 장바구니를 통한 농식품 네트워크 분석 : 상추를 중심으로\*\*\*

김은영, 위태석, 임세화, 장성원, 하지희, 채지희, 이홍진, 신혜진  
농촌진흥청 농산업경영과  
e-mail:key2022@korea.kr

## A Network Analysis of Market Basket : Focusing on Lettuce

Kim,EunYoung, Wi,TtaeSeok, Lim,Sehwa, Jang,SeongWon, Ha,Jihui,  
Chae,Jihui, Lee,HongJin, Shin,HyeJin  
Rural Development Administration

### 요약

본 논문에서는 채소 섭취 문제 해결을 위해, 특정 품목에 대한 타 식품군과 관계를 파악하여 장바구니 개선을 방안을 모색하였다. 생산량이 증가하고 있으나 대표적 육류 보완재로 고기 소비에 의존성이 큰 상추에 집중하여, 소비자가 상추를 구매할 때 고려할 가능성이 높은 품목을 도출하였다. 분석 자료로 농촌진흥청의 2010~2021년 농식품 소비자 패널조사를 사용했다. 상추를 구매한 당일의 소비 목록에서 각 농식품군 소분류 항목을 이용하여 비지도학습 기법 중의 하나인 연관규칙 분석법을 시행하였다. 상추와 상관성이 높은 품목을 중심으로 결론을 도출하고 네트워크 시각화를 통해 직관적인 분석 결과를 제시하였다. 이를 통해 장바구니 구성의 문제점을 발견하고, 소비자의 구매 시점에 있어 무의식적인 개입을 통한 현실적이고 세부적인 관점에서의 채소 섭취 문제 해결에 기여하고자 했다.

## 1. 서론

### 1.1 연구 배경 및 필요성

우리나라 국민의 채소 섭취 부족 문제가 심각하다. 질병관리청의 2021 국민건강통계에 따르면, 하루 권장량(500g)을 충족하지 못하는 만 19세 이상 성인의 비율이 71.9%로 과반수 이상이다. 하루 채소 섭취량은 '10년 322.9g에서 '21년 276.8g으로 12년 사이 46.1g 감소했다. 보건복지부는 매 5년을 주기로 식생활지침을 발표하고 있다. 그러나 1일 권장량을 준수하여 건강한 식습관을 실천하고 있는 우리나라 성인은 28.1%에 불과한 실정이다. 채소 불균형 문제를 세부적인 관점에서 해결하기 위해, 본 연구에서는 상추에 집중하여 살펴보고자 했다. 상추 생산량은 '10년 141천톤에서 '16년 86천톤으로 감소하는 추세였으나, 이후 5년간 11천톤까지 다시 증가하였다. 또한 상추의 생산액은 '21년 기준으로 업체류 전체 중에서 22%를 차지하는 주요 작물이지만 대표적인 육류의 보완재로 소비 목적이 한정적이다. 특히, 적색육인

돼지고기를 구매할 때 씬용으로 소비되는 경우가 많아 고기 소비에 크게 좌우되는 경향이 있다. 이와 관련하여 우리나라 육류 소비량은 '10년 115.0g에서 '21년 125.9g으로 10.9g이 증가했다. WHO에서 권고하는 1일 평균 적색육 섭취량(100g 미만)을 훨씬 초과한 실정으로 상추와 소비되는 품목 및 용도를 다양화할 필요가 있다. 그러나 대부분의 연구가 포괄적인 식품 유형에 대한 분석으로, 특정 품목에 대한 연구는 거의 없다. 또한 상추의 경우, 기능성 및 관능적 특성, 품질에 대한 연구가 주로 이루어져 소비 관점에서의 연구는 미흡한 실정이다.

따라서 보완관계에 있는 타 식품군과 잠재적 연관 관계를 파악한다면 상추 소비량 재고를 통한 채소 섭취 증진에도 긍정적인 효과를 기대할 수 있을 것으로 예상된다. 이에 본 연구에서는 농촌진흥청의 '10~'21년 농식품 소비자패널데이터를 활용하여, 장바구니 분석(MBA:Market Basket Analysis)을 통한 농식품 네트워크를 분석하고자 한다. 분석 결과를 통해 상추 소비와 양의 상관관계에 있는 타 식품군을 도출하여 소비자 스스로 건강한 장바구니를 구성하는데 도움을 주고 상추 소비 촉진을 통한 채소 섭취 증가에 기여하고자 한다.

\* 본 성과물은 농촌진흥청 연구사업(과제번호:PJ015127)에 의해 이루어진 것임.

\*\* 본 연구는 2023년도 농촌진흥청 전문연구원 과정 지원 사업에 의해 이루어진 것임.

### 1.2 선행연구 검토

원은송 외 1인(2020)의 연구에서는 연관규칙 분석을 통해 소비자가 구매하는 신선식품 간의 상호 연관성을 살펴보았다. 연령대와 구입장소별로 향상도 결과를 확인하고 이를 통한 마케팅 전략을 도출하였다.

윤선미 외 1인(2022)의 연구에서는 1인 가구원의 배달 및 테이크아웃 음식의 선택요인을 연결망 분석으로 살펴보았다. 각 연결망에서 중심성 분석을 진행하였으나 요인 간 관계 특성에 대한 한계를 언급했다.

A. Gupta 외 3인(2016)의 연구에서는 연관규칙 분석법의 데이터 처리 효율성을 고려하여 Apriori, ELAT 및 FP-Growth를 비교하였다. 그 결과 Apriori가 큰 데이터 셋을 처리하는데 소요되는 시간이 가장 효율적이었다.

이동원(2017)의 연구에서는 연관규칙의 확장기법으로 두 상품 간의 연결망 중심성 척도를 회귀분석의 변수로 설정하여 연관성이 나타나지 않는 상품 간의 잠재적 연관성을 도출하였다.

### 1.3 연구의 목적

본 연구에서는 장바구니 분석을 통해 상추와 상관성이 높은 품목을 도출하고 연관 품목과의 상품 구성을 제시한다. 소비자 스스로 건강하고 합리적인 농식품 소비 선택을 할 수 있도록 유도함으로써 채소 섭취 부족 문제를 해결할 수 있는 방안을 도출한다.

## 2. 자료 및 분석방법

### 2.1 분석 자료

연구의 분석 자료는 농촌진흥청의 '10년~'21년 농식품 소비자패널 데이터에서 상추를 구매한 당일의 구매데이터를 사용하였다. 이를 위해 각 소비자에게 부여된 고유 번호와 구매 일자를 기준으로 식품 분류 중 소분류에 대해 집계하였다. 패널은 1년 9회 이상 가계부를 작성해 온 1,988가구를 대상으로 했다. 식품 분류에는 대분류, 중분류, 소분류가 있는데 품목 간 연관성을 세부적으로 살펴보기 위해 469개로 분류된 소분류 데이터를 사용하여 분석을 진행하였다.

[표 1] 분석 자료 1차 가공

shop_day	panel_code	가공우유	...	흰우유
2010-01-01	PN00000001	0	...	1
2010-01-01	PN00000002	1	...	0
⋮	⋮	⋮	...	⋮
2021-12-31	PN20170466	1	...	1

### 2.2 분석방법

상추와 타 식품 간의 상호 연관성을 살펴보고 소비자에게 더 나은 구매를 유도하기 위해, 비지도 학습(Unsupervised Learning) 기법인 연관분석을 시행했다. 최적의 판단 및 예측을 가능하게 해주는 기계학습법으로 연구에 사용된 97,716건의 데이터를 처리하기에 효율적인 Apriori 알고리즘을 적용했다. 상추와 타 농식품 간의 연관 패턴 및 규칙을 한 번에 파악할 수 있는 SNA 시각화까지 분석에 적용하였다.

#### 2.2.1 연관분석(Association Analysis)

연관분석은 하나의 재화나 서비스 등을 거래하는데 포함된 항목 간의 관련성을 파악한다. 항목 사이에 존재하는 주요 특징 및 연관된 규칙을 찾아내는 탐색적 분석 방법(EDA: Exploratory Data Analysis)이다.

연관규칙(Association Rule)은 연관분석을 위해 특정 항목들 간의 관련성을 수치화시켜 나타내는 방법으로 지지도, 신뢰도, 향상도를 산출하여 연관성의 여부를 판단한다.

지지도(Support)는 두 개 이상의 품목이 동시에 거래되는 확률을 나타낸다. Support(A→B)와 Support(B←A)는 상호 대칭적으로 같은 값을 가지는 Support(A↔B)로 나타낼 수 있으며, 식은 다음과 같다.

$$Support = P(A \cap B) = \frac{\text{품목 A와 B의 동시 거래 수}}{\text{전체 거래 수}}$$

신뢰도(confidence)는 하나의 품목이 구매될 때 다른 품목이 구매되는 경우의 조건부확률이다. 지지도에서 거래 포함 비중이 낮은 경우 연관성 판단에 어려움이 있기 때문에 품목A가 포함된 거래 중에서 품목B를 포함한 거래 비율을 제공한다.

$$Confidence(A \rightarrow B) = P(B|A) = \frac{Support}{\text{품목 A를 포함한 거래 수}}$$

향상도(Lift)는 상품 간의 독립성 여부를 판단하는 척도이다. 지지도 또는 신뢰도에서 우연히 연관성이 높게 측정된 부분을 보완하기 위해 사용한다.

$$Lift(A \rightarrow B) = \frac{Confidence}{\text{품목 B를 포함한 거래율}}$$

#### 2.2.2 연결망분석(Social Network Analysis)

연결망은 개인이나 조직과 같은 사회적 개체 간의 상호작용 구조를 나타낸다. 이러한 개체간의 관계를 이해하려는 시도로 노드(node), 링크(Link)로 구성된

네트워크의 특성을 파악하기 위해, 연결정도중심성, 근접중심성, 매개중심성 등을 활용한다. 이를 통해 정량화된 연결망에 그래프 이론을 도입함으로써 사용자에게 관계의 성격 및 특징을 직관적이고 용이하게 파악할 수 있도록 한다. 그러나 대부분의 연결망 시각화가 디자이너의 주관적 판단에 의해 네트워크 레이아웃의 형태, 노드의 위치, 형태, 크기가 결정되어 관계의 성격을 명확하게 보여주지 못하는 경우가 많았다(장선희, 2009). 이에 본 연구에서는 연관분석의 결과를 통해 구조적인 데이터 시각화로 분석 결과를 전달하고자 한다.

### 3. 분석결과

#### 3.1 연관분석 결과

상추를 중심으로 향상도가 높은 타 품목과의 연관 분석 결과를 살펴보면 다음과 같다. 상추는 소주, 맥주, 탄산음료 등의 알콜성 및 비알콜성 음료와 향상도가 가장 높은 것으로 나타났다. 상추는 음료와 상관성이 높으며, 특히 맥주를 구매할 때 소주와 상추를 구매할 확률이 높았다. 다음으로 음료와 가공식품과의 상관성이 가장 높았다. 햄, 수산물가공식품, 과자류, 봉지라면과 같은 가공식품과 향상도가 높은 것으로 나타났으며, 그중에서 상추와 햄, 수산물가공식품의 조합으로 구매하는 경우가 많았고, 상추와 봉지라면, 과자류의 조합으로 구매하는 경우가 많았다. 농식품 구매 시점에서 장바구니 구성의 개선이 필요함을 확인할 수 있다. 이에 본 연구의 목적은 소비자의 장바구니 개선을 통한 채소류 섭취량 재고이기 때문에, 음료 및 가공식품을 제외한 품목과의 연관성을 살펴보았다.

[표 1] 상추와 타 식품의 연관분석 결과

antecedents → consequents	support	confidence	lift
맥주,*기타상추 → 소주	0.014	0.437	5.615
소주 → 맥주,기타상추	0.014	0.181	5.615
기타상추,소주 → 맥주	0.014	0.339	5.280
맥주 → 기타상추,소주	0.014	0.219	5.280
탄산음료,기타상추 → 소주	0.010	0.271	3.478
소주 → 탄산음료,기타상추	0.010	0.129	3.478
기타상추,소주 → 탄산음료	0.010	0.242	3.306
탄산음료 → 기타상추,소주	0.010	0.137	3.306
기타상추,햄 → 기타수산물가공식품**	0.013	0.447	3.290
기타수산물가공식품 → 기타상추,햄	0.013	0.098	3.290
기타수산물가공식품,기타상추 → 햄	0.013	0.188	3.048
햄 → 기타수산물가공식품,기타상추	0.013	0.217	3.048
과자류, 적상추 → 봉지라면	0.012	0.263	2.987
봉지라면 → 과자류, 적상추	0.012	0.133	2.987
봉지라면,적상추 → 과자류	0.012	0.456	2.869

분석 결과는 두부, 콩나물, 대파, 양파와 상관성이 높은 것으로 나타났다. 식품공전에 따르면, 두부는 두류를 제조·가공한 식품이나 두류분을 포함하여 두류가 100%인 식품으로 다른 첨가물이 거의 포함되어 있지 않은 식품이다. 또한 단백질 함량이 50%, 지방 33%를 함유하여 고기류의 대체제이자 상추와 보완재로 구성하기 좋은 식품으로 예상된다. 두부와 함께 콩나물을 구매하는 것으로 나타났는데 콩나물 역시 단백질 함량이 높은 식품이다. 따라서 상추와 두부, 콩나물을 인접한 매대에 진열하는 것이 장바구니 개선에 긍정적인 영향을 미칠 것으로 예상된다. 눈에 띄는 것은 상추와 탄산음료를 구매할 때, 삼겹살을 포함할 확률이 37%였으나, 삼겹살을 구매할 때 상추와 탄산음료를 구매할 확률은 6%였다. 이는 소비자가 삼겹살 구매 시, 상추와 탄산음료를 동시에 구매하거나 삼겹살만 구매할 확률이 높다고 볼 수 있다.

[표 2] 상추와 타 식품의 연관분석 결과

antecedents → consequents	support	confidence	lift
두부,기타상추 → 콩나물	0.023	0.223	1.952
콩나물 → 두부,기타상추	0.023	0.200	1.952
두부,적상추 → 흰우유	0.011	0.281	1.943
흰우유 → 두부,적상추	0.011	0.075	1.943
기타상추,대파 → 일반양파	0.010	0.162	1.907
일반양파 → 기타상추,대파	0.010	0.123	1.907
일반양파,기타상추 → 대파	0.010	0.194	1.902
대파 → 일반양파,기타상추	0.010	0.102	1.902
기타상추,콩나물 → 두부	0.023	0.333	1.876
두부 → 기타상추,콩나물	0.023	0.129	1.876
탄산음료,기타상추 → 삼겹살	0.014	0.373	1.831
삼겹살 → 탄산음료,기타상추	0.014	0.068	1.831

#### 3.2.2 연결망분석 시각화(SNA)

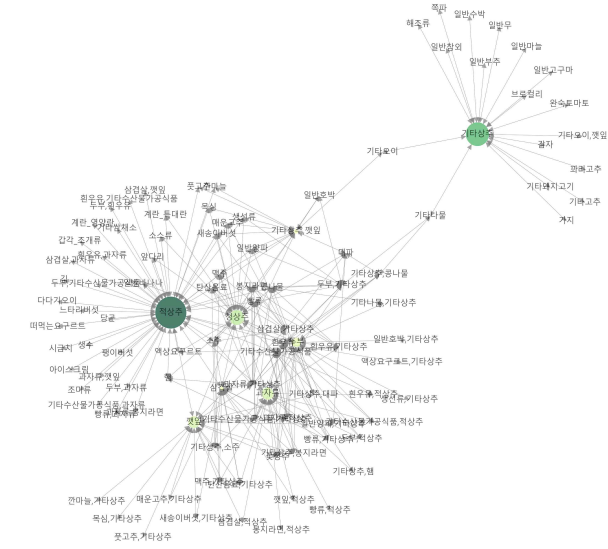
분석 결과, 연결 중심성(deg\_cen)에서 적상추와 청상추가 네트워크의 주요 품목인 것으로 나타났다. 근접중심성(clo\_cen)은 연결 중심성에서 나타난 품목들과 근접하게 연결되어 영향 정도를 높이는 품목으로 두부가 3위였다. 매개중심성(bet\_cen)에서 상추와 타 식품군 사이에 매개 역할을 하는 품목은 두부가 5위로 나타났으며, 그 외 나물, 깻잎 등도 매개역할을 하는 것으로 나타났다.

\* 적상추, 청상추, 꽃상추를 제외한 모든 상추

\*\* 참치캔, 부각, 쥐포, 핫바, 맛살, 소시지, 김자반 등

[표 3] 상추와 타 식품의 중심성 분석 결과

detail	deg_cen	detail	clo_cen	detail	bet_cen
적상추	0.4712	적상추	0.4883	적상추	0.5012
청상추	0.2404	과자류	0.4370	기타상추	0.2858
과자류	0.1923	두부	0.4262	기타나물	0.1646
깻잎	0.1827	삼겹살	0.4160	기타상추, 깻잎	0.1334
기타상추	0.1731	기타나물, 기타식품	0.4127	두부, 기타상추	0.1295



[그림 1] 상추와 타 식품군과의 연결망

4. 결론 및 시사점

연관 분석은 마케팅의 관점에서 소매 유형별 상품 구성 및 추천 시스템 알고리즘으로 사용되어 왔다. 이 연구에서는 채소 섭취 재고를 위해, 소비 시점에서 장바구니를 개선하여 소비자의 합리적인 구매 결정을 유도하고자 했다. 또한 대부분의 연구는 확장된 개념의 식품 유형에 관한 연구였다면, 세부적인 관점에서 특정 품목인 상추에 집중하여 분석을 진행했다. 이를 위해 상추와 동시 구매 확률이 높은 품목을 분석하여 상추 소비를 촉진하고 문제 해결 방안을 도출하고자 했다. 그 결과 소비자들은 상추를 구매할 때 알콜성 및 비알콜성 음료와 가공식품을 구매할 확률이 가장 높았다. 또한 상추 구매 시 음료 및 가공식품 구매 확률보다 음료 및 가공식품 구매 시 상추 구매 확률이 높은 것으로 나타났다. 장바구니 내에서 음료 및 가공식품의 구매가 상당히 높은 비중을 차지하며, 식습관에 대한 비합리적인 농식품 장바구니를 구성하고 있다는 것을 알 수 있었다. 상추 구

매 시 건강한 장바구니에 도움이 될 품목은 두부가 가장 적합하였다. 중심성 분석 결과, 상추와 밀접하면서 타 식품 간 매개역할을 하는 품목은 두부가 삼겹살보다 높아 향후 상추는 두부와 양의 상관성을 띄는 보완관계로 성장할 가능성이 있다. 따라서 상추와 두부를 밀접한 매대에 진열하고 판매한다면 소비자 시점에서 두 품목에 대한 동시 선택으로 장바구니 개선을 통한 채소 섭취 권장량 충족과 육류 보완재로 소비되는 경향을 두부로 분산하여 육류 섭취 과다 문제를 감소시키는데도 실제적인 도움이 될 것으로 예상된다.

한편, 본 연구에서는 상관관계 외에 인과관계를 규명하지 못하였다. 따라서 후속 연구에서는 QAP 다중 회귀 분석 모델 등을 사용하여 보완하고자 한다.

참고문헌

- [1] 원은송, 김성용, “연관규칙 기법을 활용한 소비자의 신선식품 구매 패턴 분석”, 농업생명과학연구, 제 54권 4호, pp. 111-122, 7월, 2020년.
- [2] 윤선미, 박득희, “1인 가구원의 배달 및 테이크아웃 음식 선택요인 연결망 연구: 사회연결망 분석을 적용하여”, 호텔관광연구, 제 24권 95호, pp. 123-137, 12월, 2022년.
- [4] Ambuj Gupta, Sudhakar Hannah, Kappor Shivam and Anand Shibangi, “Performance Comparison of Apriori, Eclat and FP-Growth Algorithm for Association Rule Learning”, International Journal of Advanced Science Computing and Engineering, vol2, No3, 2016.
- [5] 김진성, “빅데이터 분석을 위한 R 프로그래밍”, 11월, 2020년.
- [6] 이동원, “사회연결망 분석을 활용한 연관규칙 확장기법”, 한국지능정보시스템학회, 제 23권 4호, pp. 111-126, 12월, 2017년.
- [7] 장선희, 장석현, “사회연결망 영향력 시각화를 위한 프레임워크”, 한국멀티미디어학회, 제 12권 1호, pp. 139-146, 1월, 2009년.