

보건복지 ISSUE & FOCUS

KIHASA
한국보건사회연구원
Korea Institute for
Health and Social Affairs
www.kihasa.re.kr

제 312호 (2016-12)
발행일 2016. 04. 18.
ISSN 2092-7117

발행인 김상호 발행처 한국보건사회연구원 (30147)세종시 시청대로 370 세종국책연구단지 사회정책동(1층~5층) TEL 044)287-8000 FAX 044)287-8052

소셜 빅데이터를 활용한 식품안전에 관한 감성분석



송태민
정보통계연구실 연구위원

- 식품안전 감성분석 결과, 식품안전 인식에 대한 긍정감정은 37.0%, 보통감정은 39.8%, 부정감정은 22.9%로 나타남.
- 온라인 문서에 '부정불량식품'이 있을 경우 식품안전에 대한 부정감정은 이전의 38.0%에서 52.4%로 크게 증가한 반면, 긍정감정은 62.0%에서 47.6%로 감소함.
- 식품안전 체감도에 대한 대국민 조사와 함께 소셜 미디어에서 수집된 빅데이터의 활용과 분석을 병행할 경우, 정부의 식품안전 정책에 대한 예측 및 평가에 대한 신뢰성은 더욱 높아질 것임.

1. 서론

■ 식품안전을 둘러싼 환경이 지속적으로 변화하면서 다양한 원인으로 인한 식품사고나 문제는 식품안전에 대한 국민의 불안을 증대시키고 있음.¹⁾

- 각종 가공식품이나 인스턴트식품에 대한 소비자들의 수요 증가와 함께 수입식품으로 인한 식품안전 위험에 대한 우려가 팽배해짐.
- 식품안전의 문제는 세균 또는 바이러스로 인한 식중독 외에 허가되지 않은 식품첨가물의 사용이나 농약, 동물의약품, 방사능 유출로 인한 식품오염이 이슈화되고 있음.

- 식품안전 사고에 대한 효과적인 대응을 위해서는 국민과 행정당국의 쌍방커뮤니케이션이 중요하며, 식품안전에 대한 위험을 예측하여 적시대응 체계를 구축할 필요가 있음.

■ 최근 소셜 네트워크 서비스(SNS)의 규모와 영향력이 점차 증대됨에 따라, 개인의 커뮤니케이션 수단으로서의 SNS는 긴급 상황과 위기대응에 결정적인 요소로 진화함.

- 트위터, 블로그, 카페, 게시판, 뉴스채널 등 온라인 매체에 남긴 정치, 경제, 사회, 문화에 대한 메시지가 그 시대의 감성과 정서를 파악할 수 있는 원천으로 등장함에 따라, 소셜 빅데이터 분석을 통해 정책수요나 사회위험에 대한 예측이 가능하게 됨.

1) 김정선 외(2015). 국민건강과 안전을 위한 식품안전 전략구축 방안. 한국보건사회연구원.

- 소셜 빅데이터 분석은 방대한 양의 데이터를 활용하여 다양한 참여자의 생각과 의견을 확인할 수 있기 때문에, 기존의 오프라인 조사와 병행하면 사회적 이슈에 대한 현황을 보다 유용하게 파악할 수 있음.
- 소셜 빅데이터의 분석은 사용자가 남긴 문서의 의미를 분석하는 것으로 자연어 처리기술인 주제분석(text mining)과 감성분석 기술인 오피니언 마이닝(opinion mining)을 실시한 후, 통계 분석(statistical analysis)과 데이터마이닝 분석(data mining analysis)을 실시함.

2. 식품안전 인식수준 예측

가. 분석대상 및 분석방법

- 본 연구에서는 257개의 온라인 뉴스사이트, 18개의 게시판 및 주요 커뮤니티, 1개의 SNS(트위터), 4개의 블로그, 2개의 카페 등 총 282개의 온라인 채널을 통해 수집 가능한 텍스트(text) 기반의 웹문서(버즈)를 소셜 빅데이터로 정의함.
- 식품안전 토픽(topic)은 모든 관련 문서를 수집하기 위해 '식품안전'이라는 토픽을 사용하였으며, 같은 의미로 사용되는 토픽 유사어로는 '식품위험, 음식안전, 먹거리안전, 먹거리위험, 음식위험, 위험식품, 먹거리사고, 먹거리안전사고, 식품사고, 식품위생, 불량식품, 부정식품' 용어를 사용함. '식품건조기, 교통사고, 안전산업기사, 안전기사, 안전거래, 식품조리과, 식품영양학과 등'은 불용어(stop-words)로 사용하여 수집/분석에서 제외함.
- 소셜 빅데이터의 수집은 2011. 1. 1.~2015. 6. 30.의 기간에 해당 채널에서 요일, 주말, 휴일을 고려하지 않고 매 시간 단위로 수집이 이루어졌고²⁾, 수집된 총 2,077,112건의 텍스트 문서 중 감정을 표현한 434,850건의 문서를 본 연구의 분석에 포함함.

- 식품안전 인식을 예측하는 가장 효율적인 예측모형을 구축하기 위해 데이터마이닝의 연관규칙과 의사결정트리 분석을 사용함.
- 연관규칙의 분석 알고리즘은 선형적 규칙(apriori principle)을 사용하였고, 의사결정트리 형성을 위한 분석 알고리즘은 CHAID(Chi-squared Automatic Interaction Detection)를 사용함. 그리고 데이터마이닝 분석은 R 3.2.1과 IBM SPSS 23.0을 사용함.

나. 분석결과

- 식품안전 관련 감성분석과 주제분석³⁾
 - 식품안전 관련 감정은 감성어 사전을 개발하여 긍정감정은 '4대약 근절하다, HACCP 좋다, 가공식품 안전하다, 길거리음식 안전하다 등', 보통감정은 '4대약 알다, MSG 먹다, 가공식품 다양하다, 길거리음식 찾다 등', 부정감정은 '4대약 웃기다, MSG 위험하다, 가공식품 불안하다, 길거리음식 위험하다 등'으로 구분하여 감성분석을 실시함.
 - 식품안전에 대한 식품위해요소는 주제분석을 통하여 'GMO, 방사능물질, 방사선, 중금속, 세균 등' 22개 요인으로 그룹화 하였고, 식품명은 '학교식품, 수산물, 과자류, 유제품, 알가공품 등' 19개 요인으로 그룹화 함.
- 식품안전 관련 연도별 감정(체감도) 변화
 - 식품안전 감성분석 결과, 긍정감정(안전)이 37.3%, 보통감정이 39.8%, 부정감정(불안)이 22.9%로 나타나, 식품안전에 대해 긍정적인 문서가 부정적인 문서보다 많음.
 - 기존의 조사(통계청 등)와 비교해 볼 때, 식품안전에 대한 긍정적인 감정은 차이가 비교적 컸으며, 부정적 감정은 2012년 33.4%, 2013년 27.9%, 2014년 26.2%인데 비해 소셜 분석에서는 각각 22.8%, 23.0%, 22.8%로 나타남<표1 참조>.

2) 본 소셜 빅데이터의 수집은 '김정선외(2015). 국민건강과 안전을 위한 식품안전 전략구축 방안. 한국보건사회연구원 연구의 일환으로 (주)SK텔레콤 스마트인사이트'에서 수행함

3) 식품안전 관련 위해요소는 크게 제조공정, 관리방식, 취급방식에 따라 발생하는 외부 위해요소(첨가물, GMO 등)와 식품자체에서 발생하는 내부 위해요소(농약, 발암물질 등)로 분류하였으며, 식품안전 관련 식품명은 가공식품(소시지 등 가공식품, 유제품, 건강기능식품, 학교식품, 패스트푸드, 야식류 등)과 비가공식품(농산물, 수산물, 육류 등)으로 분류하였음. 식품안전의 주제분석은 이화여자대학교 식품영양학과 조미숙 교수 연구팀에서 국내외 논문 게재를 위해 개발한 온톨로지(식품안전 관련 소셜 빅데이터 온톨로지)를 수정하여 사용함.

〈표 1〉 식품안전 감정(체감도)에 대한 설문조사와 SNS 조사 결과 비교

(단위: %)

연도	긍정		보통		부정		$\chi^2(p)^{3)}$
	기존 조사	소셜	기존 조사	소셜	기존 조사	소셜	
2011년		38.3		38.2		23.5	379.994 (0.000)
2012년	16.1 ¹⁾	36.4	50.5 ¹⁾	40.7	33.4 ¹⁾	22.8	
2013년	18.2 ¹⁾	35.9	54.0 ¹⁾	48.4	27.9 ¹⁾	23.0	
2014년	25.0 ²⁾	37.6	48.9 ²⁾	39.6	26.2 ²⁾	22.8	
2015년		39.0		39.0		22.0	

자료: 김정선 외(2015). 국민건강과 안전을 위한 식품안전 전략구축 방안. 한국보건사회연구원. p100.

주: 1) 통계청(2012). 2012 통계청 사회조사 (보도자료)

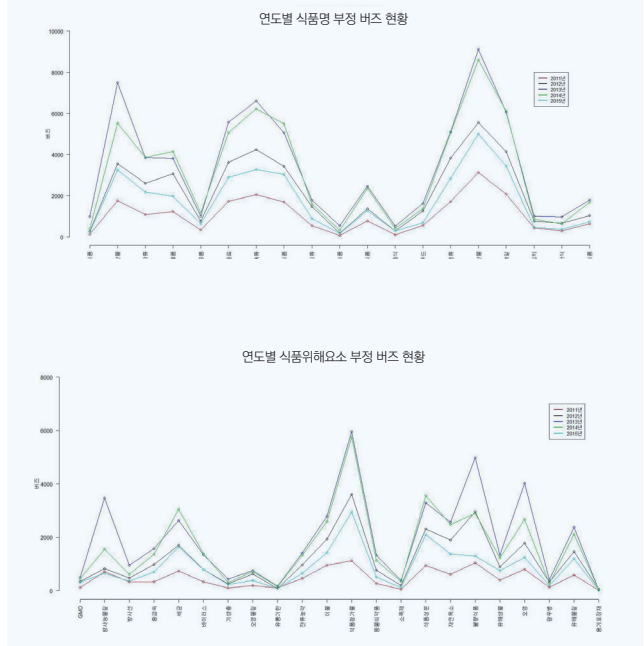
2) 정기혜 등(2014). 2014년도 식품안전체감도 조사, 국무총리실·한국보건사회연구원 (상·하반기 평균), p.182.

3) 소셜 빅데이터의 분석 결과

■ 식품안전 관련 식품명과 식품위해요소 부정감정 온라인 문서(버즈) 현황

- 식품명에 대한 부정감정은 농산물, 육류, 과일, 수산물, 음료, 가공식품 등의 순으로 높은 것으로 나타남.
- 식품위해요소에 대한 부정감정은 식품첨가물, 불량식품, 식품성분, 오염, 세균, 이물질 등의 순으로 높은 것으로 나타남(그림1 참조).

〔그림 1〕 식품안전 관련 식품명과 식품위해요소 부정감정



■ 식품안전 위해요소에 대한 연관성 분석

- 식품안전 위해요소에 대한 연관성 예측에서 가장 신뢰도가 높은 연관규칙은 {식품첨가물, 소독제, 식품성분, 유해물질 유형}={부정}이며, 다섯 개 변인의 연관성은 지지도 0.001, 신뢰도 0.870, 향상도 2.286으로 나타남. 온라인 문서에서 '식품첨가물, 소독제, 식품성분, 유해물질 유형'이 언급되면 식품안전의 감정이 부정적일 확률이 약 87.0%이며, '식품첨가물, 소독제, 식품성분, 유해물질 유형'이 언급되지 않은 버즈보다 식품안전에 대한 인식이 부정적일 확률이 약 2.29배 높아지는 것으로 나타남.

〈표 2〉 식품안전 관련 위해요소 연관규칙

규칙	지지도	신뢰도	향상도
{식품첨가물,소독제,식품성분,유해물질 유형} => {부정}	0.001197965	0.8693694	2.285853
{식품첨가물,소독제,유해물질 유형} => {부정}	0.001654185	0.8446910	2.220965
{중금속,식품첨가물,동물용의약품,자연독소} => {부정}	0.002340067	0.8322296	2.188200
{식품첨가물,소독제,식품성분} => {부정}	0.001477284	0.8292683	2.180414
{GMO,세균,식품첨가물} => {부정}	0.001055203	0.8292683	2.180414
{소독제,식품성분,유해물질 유형} => {부정}	0.001219690	0.8170478	2.148283
{중금속,식품첨가물,동물용의약품} => {부정}	0.003131469	0.7995246	2.102208
{GMO,식품첨가물,식품유해생물} => {부정}	0.001033478	0.7798595	2.050502
{소독제,유해물질 유형} => {부정}	0.001790741	0.7786775	2.047394
{식품첨가물,소독제} => {부정}	0.002597661	0.7735675	2.033959

규칙	지지도	신뢰도	향상도
[중금속, 동물용의약품, 자연독소] => [부정]	0.002914221	0.7696721	2.023717
[세균, 광우병] => [부정]	0.001048996	0.7664399	2.015218
[소독제, 식품성분] => [부정]	0.001827983	0.7503185	1.972830
[GMO, 광우병] => [부정]	0.001173137	0.7382812	1.941180
[중금속, 동물용의약품] => [부정]	0.003972527	0.7223476	1.899285
[GMO, 잔류농약, 식품첨가물] => [부정]	0.001297279	0.7219344	1.898199
[GMO, 잔류농약, 식품유해생물] => [부정]	0.001064513	0.7014315	1.844290
[방사능물질, 식품첨가물, 오염] => [부정]	0.001371763	0.6949686	1.827297
[오염물질, 자연독소, 오염, 유해물질 유형] => [부정]	0.001483491	0.6937591	1.824117
[방사능물질, 방사선오염] => [부정]	0.003193540	0.6887550	1.810959
[식품첨가물, 부정불량식품, 유해물질 유형] => [부정]	0.001964539	0.6872964	1.807124
[오염물질, 자연독소, 오염] => [부정]	0.001840398	0.6871379	1.806707

■ 식품안전 인식수준에 대한 식품위해요소 예측모형

- 식품안전 관련 식품위해요소가 식품안전의 인식수준 예측모형에 미치는 영향은 '부정불량식품'이 가장 큰 것으로 나타남.
 - 온라인 문서에서 '부정불량식품'이 있을 경우 식품안전의 부정감정은 이전의 38.0%에서 52.4%로 크게 증가한 반면, 긍정감정은 이전의 62.0%에서 47.6%로 감소함.
 - 온라인 문서에 '부정불량식품'이 있고 '식품첨가물'이 있는 경우 식품안전의 부정감정은 이전의 52.4%에서 63.9%로 증가한 반면, 긍정감정은 이전의 47.6%에서 36.1%로 감소함.

■ 식품명에 대한 부정감정은 농산물, 육류, 과일, 수산물, 음료, 가공식품 등의 순으로 높은 것으로 나타났으며, 식품위해요소에 대한 부정감정은 식품첨가물, 불량식품, 식품성분, 오염, 세균, 이물질 등의 순으로 높은 것으로 나타남.

■ 식품안전 인식수준 예측에서 온라인 문서에 '부정불량식품'이 있을 경우 식품안전의 부정감정은 이전의 38.0%에서 52.4%로 크게 증가한 반면, 긍정감정은 이전의 62.0%에서 47.6%로 감소함.

■ 온라인상에 식품안전에 대한 불안한 징후가 예측되면 소셜 빅데이터 분석 모델을 활용한 불안수준 모니터링 체계를 구축하고(김정선 외, 2015), 잘못된 정보를 바탕으로 형성된 부정적인 정보에 대한 정확한 정보를 제공하고 확산시키는 '위해정보전달(risk communication)' 체계가 구축되어야 함.

■ 식품안전 체감도에 대한 대국민 조사와 더불어 소셜 미디어에서 수집된 빅데이터의 활용과 분석을 병행할 경우, 정부의 식품안전 정책에 대한 예측 및 평가에 대한 신뢰성은 더욱 높아질 것임.

3. 요약 및 결론

■ 식품안전 감성분석 결과 식품안전에 대한 긍정감정은 37.0%, 보통감정은 39.8%, 부정감정은 22.9%로 나타남.