

폭염에 의한 축산폐사와 뉴스 빅데이터의 상관관계 분석*

박종철** · 한국진*** · 채여라****

Correlation Analysis between Livestock Mortality Caused by Heat Wave and News Big Data*

Jongchul Park** · Kukjin Han*** · Yeora Chae****

요약 : 소셜 빅데이터는 재난을 조기 탐지하는 정보의 원천이 될 수 있으며, 재난의 공간적 분포를 이해하기 위한 새로운 가능성을 내포하고 있다. 이를 위해서는 뉴스 빅데이터로부터 수집된 정보와 실제 사건의 관계에 대한 이해가 선행되어야 한다. 본 연구의 목적은 뉴스 빅데이터의 분석 결과와 폭염에 의한 가축 폐사와의 비교를 통해 두 자료의 관계에 대한 이해를 증진시키는 것이다. 가축 폐사가 증가하는 기온 구간에서 축산 피해 관련 뉴스는 다른 시기에 비해 두 배 이상 증가하였다. 하지만 뉴스 건수의 정점은 가축 폐사의 정점으로부터 약 6일 후에 나타나고 있었다. 가축 폐사가 증가하는 기온 구간에서 뉴스의 주요어는 '폐사'이었다. 7월 중순 이전의 뉴스에서 주요어는 '대응', '예방'이었고, 7월 중순부터 8월 중순에는 '폐사', 8월 중순 이후에는 '물가'가 주요어를 이루고 있었다. 사회적 이슈에 의해 특정 주요어의 빈도가 높아지기도 하지만 '폐사'라는 주요어는 대체로 실제 폐사가 집중되는 기온 구간 및 시기에 등장하고 있었다.

주요어 : 기사, 빅데이터, 폭염, 폐사, 재난

Abstract : Social big data can be a source of information for early detection of disasters. Furthermore, it contains new possibilities for understanding the spatial distribution of disasters. Understanding the relationship between information obtained from news big data and actual events is essential to do this. The purpose of this study is to improve the understanding of the relationship between the two data by comparing the results of the analysis of news big data and livestock mortality caused by heat waves. The number of news was doubled during the period livestock mortality increased. However, the number of news reached the peak after six days the livestock mortality reached the peak. In the temperature range where livestock mortality increased, the main keyword of the news was 'mortality'. In the news before mid-July, the main keywords were 'response' and 'prevention', and the main keyword was 'mortality' from mid-July to mid-August. Since mid-August, 'price' was the main keyword. Although the frequency of some key words is affected by social issues, the key word 'mortality' appeared mostly in temperature ranges and periods of actual mortality.

Key Words : News, Big data, Heat wave, Mortality, Hazard

*이 연구는 기상청 <자연재해대응 영향예보 생산기술 개발>(KMI2018-01410)의 지원으로 수행되었음.

**한국환경정책·평가연구원 초빙연구원(Invited Research Fellow, Korea Environment Institute, jcpark@kei.re.kr)

***한국환경정책·평가연구원 선임전문원(Technician, Korea Environment Institute, kjhan@kei.re.kr)

****한국환경정책·평가연구원 선임연구위원(Senior Research Fellow, Korea Environment Institute, yrchae@kei.re.kr)

I. 서론

최근 폭염은 국가 차원의 재난대응이 요구되고 있으며 2018년부터 법정 '자연재난'에 포함되었다. 폭염이 사람 뿐 아니라 동·식물에 다양한 악영향을 미친다는 것은 잘 알려져 있다(채여라 등, 2018). 농촌진흥청(2018)에 의하면 폭염에 의해 소, 돼지, 닭 등 가축의 체온이 상승하고, 음수량이 증가하며 사료 섭취량은 감소한다. 기온이 일정 수준, 즉 임계기온을 넘게 되면 가축에게서 성장저하 또는 대사성 질병과 식중독 등이 발생하고 심한 경우 가축이 폐사에 이르기까지 한다. 폭염이 가축에게 미치는 악영향은 생리특성 연구를 통해 제시되거나(하태환 등, 2018), 폭염 대비 가이드라인으로 공개되고 있다(국립축산과학원, 2016; 농촌진흥청, 2018). 하지만 폭염이 가축에 미치는 영향은 축사 시설과 사료, 관리 방안 등에 의해 다르기 때문에 축종별 생리특성만으로 폭염 피해를 예측하는데 한계가 있다. 따라서 폭염 재난 관리 측면에서 실제 피해가 발생했을 때 피해를 조기에 탐지하고 재난 상황에 빠르게 대처하는 것이 중요하다.

피해를 조기에 탐지하고 재난 상황을 전파하는 것은 재난의 방지 및 관리에서 중요한 역할을 한다(한국기후변화대응전략연구소, 2013). 최근 비정형 빅데이터로서의 뉴스는 가까운 미래에 발생 가능한 위험요인의 포괄적 약신호(weak signal)를 효율적으로 획득할 수 있는 정보의 원천으로서 주목받고 있다(국립재난안전연구원, 2017). 이공희(2017)에 의하면 2015년 기준 비정규직을 포함한 기자의 수는 25,951명이며 기자 당 연 평균 50건의 기사를 작성한다고 가정했을 때 이들에 의해 생성되는 뉴스 기사는 100만 건을 넘는다. 뉴스는 텍스트, 사진, 그래프 등이 중심이 된 비정형 데이터이며 실시간으로 작성되고 축적되므로 빅데이터의 일반적인 특성인 3V(Volume, Variety, Velocity)를 모두 가지고 있다. 통계자료 등을 활용한 전통적인 방법의 재난 모니터링은 정확한 현황을 파악할 수 있다는 점에서는 장점이 있지만 자료가 수집되고 확정되기까지 상당한 시일이 걸리기 때문에 조기 탐지와 예측에 활용하는 데 한계가 있다. 이에 반하여 뉴스는 빠른 정보 수집과 지속적인 갱신이 가능하고, 주요어 변화분석만으로 특정 현상의 역추적이 가능하기 때문에 재난 분야에서 주목받고 있다. 특히 기후와 사회환경 등의 변화에 의해 재난이 복잡화, 대형화, 신종화 되면서 재난에 대한 예측과 전망에 대한 중요

성이 강조되면서 뉴스와 같은 비정형 빅데이터의 활용에 대한 관심이 높아지고 있다(국립재난안전연구원, 2017; 정우석 등, 2019).

최선화(2014)는 도시 내수침수 전조를 조기에 감지하고 위험지역을 분석하기 위한 방안으로 침수예상도 등 기존의 정형데이터와 소셜네트워크 서비스, 뉴스 등 비정형데이터를 연계 활용하는 모델을 제안한 바 있다. 노병준 등(2016)은 뉴스의 키워드를 분석하여 구제역의 사회적 파급효과를 분석하였다. 국립재난안전연구원(2017)은 재난 뉴스의 내용을 분석하여 재난 주제별로 분류하는 기술을 개발하고 조류 독감, 포항 지진, 경주 지진을 사례로 재난의 경과에 따른 뉴스 빈도와 국민 정서의 변화를 분석하였다. Ihm *et al.*(2017)은 식품안전 사고(food hazard event)를 조기 탐지하기 위한 방안으로 온라인 뉴스 및 소셜 미디어를 자동으로 수집하고, 분석하는 시스템을 제안하였으며, 정우석 등(2019)은 대형복합 재난 피해확산예측 시스템을 구성하는 한 요소로 뉴스를 반영하였다.

재난 부문에서 뉴스를 활용한 몇몇 사례들을 살펴보면 폭염과 관련된 연구 사례는 찾아보기 어렵다. 2018년에 폭염이 자연재난으로 분류되었기 때문에 아직까지 관련 연구가 부족하기 때문이다. 아울러 뉴스 기사의 증감과 실제 피해와의 상관관계, 예를 들어 뉴스와 실제 피해와의 시기적 일치, 실제 피해가 집중되는 시기에 등장하는 주요 키워드 등에 대해 일별로 분석한 사례는 찾아보기 어려웠다. 기존 연구에서 노병준 등(2016)은 뉴스 증감과 구제역 발생 추이를 월 단위로 비교하여 일별로 생산되는 뉴스의 특성에 대한 정보를 제공하는 데 한계가 있었다. 국립재난안전연구원(2017)은 뉴스 빈도와 조류독감 사망자 수의 관계를 국외 사례를 토대로 제시하여 국내 뉴스의 특성을 제시하는 데 한계가 있었다.

뉴스 기사와 실제 피해의 상관관계를 밝히는 것은 뉴스를 재난의 포괄적 약신호를 탐색하는 정보의 원천으로 활용하는 것에 대한 과학적 근거가 될 수 있다. 이 정보는 또한 뉴스를 이용한 모니터링 시스템 설계 시 피해 탐지를 위한 주요 키워드의 적정성을 판단하는 기준이 되며 뉴스의 빈도 변화를 통해 피해 규모의 정점을 추정하는 논리적 근거가 될 수 있다. 모니터링의 특성상 일 단위의 상관관계는 특히 중요한 정보가 된다. 그럼에도 불구하고 선행 연구들에서는 이에 대해 충분한 연구가 이루어지지 않았다.

지리학은 자연재해를 지도화하여 자연재해 발생과 피해의 공간적 분포에 대한 이해를 증진시킴으로써 자연재해로 인한 피해를 경감하는데 오랜 기간 기여해 왔다 (Rosenfeld, 1998; 장은미, 2006; 장동호, 2007; 오성렬·이기하, 2014; 김동우 등, 2017; 알 마문 등, 2017; 박준범·남궁미, 2019). 비정형 빅데이터에서 재해와 관련된 키워드를 추출하면 재해의 공간적 분포를 이해하는 데 활용할 수 있는 가능성이 있다(임현택 등, 2016). 하지만 기존 국내 연구에서 뉴스의 공간적 분포가 실제 피해의 공간적 분포와는 어떤 관계에 있는지에 대해 밝힌 연구는 많지 않다.

비정형 데이터는 또한 재난의 인식 및 조기 경고, 대응, 재난 후 평가에 유용한 요소들을 포함하고 있고 지속적으로 업데이트되기 때문에 재해에 대한 중요한 정보의 원천이 될 수 있다(Tarasconi *et al.*, 2017). 소셜 빅데이터와 텍스트 마이닝을 활용한 이슈 분석 방법은 분석의 자동화와 지속적인 활용이 가능하다는 측면에서 장점이 있다(진대용 등, 2018). 하지만 비정형 데이터를 통해 수집된 정보와 실제 피해의 관계에 대한 이해가 부족하면, 재생산성과 같은 비정형 데이터의 불확실성으로 인해 그 분석 결과를 대국민 서비스 등 정책적으로 활용하기 어려운 것이 현실이다. 따라서 이러한 한계를 극복하고 비정형 데이터를 활용하여 자연재해의 공간적 분포를 이해하려는 연구는 지리학에 도전이 될 수 있으며 재난 대응 및 관리에서 중요한 연구이다.

본 연구는 폭염에 의한 가축 폐사와 축산 피해를 보도한 뉴스와의 관계에 대하여 분석하였다. 본 연구는 크게 세 가지 측면에서 기존 연구와 차별성을 갖는다. 첫째, 기존 연구에서 다루지 않았던 폭염 이슈를 다루었다. 둘째, 국내 몇몇 연구들이 시도하였지만 충분하지 않았던 뉴스와 실제 피해와의 사례를 일 단위에서 비교하였다. 셋째, 실제 피해의 공간적 분포와 뉴스의 공간적 분포를 비교하였다.

이를 위해 본 연구에서는 폭염에 의해 폐사한 가축의 수와 뉴스 건수의 증감 패턴을 비교·분석하였고, 가축 폐사가 집중된 기온과 시기, 그리고 그 시기를 전·후하여 나타나는 뉴스의 주요어 변화를 분석하였다. 또한 뉴스에 포함된 피해 지역을 지도화하고 실제 피해 발생 지역 및 주요 사육지역과 비교하였다.

II. 연구자료 및 방법

1. 연구자료

본 연구는 2015~2017년 5~9월의 기상 자료와 뉴스, 가축 폐사 자료를 이용하여 연구를 수행하였다. 기상 자료는 '기상자료개방포털'에서 제공하는 종관기상관측자료의 일 최고기온을 이용하였으며, 뉴스 자료는 한국언론진흥재단에서 제공하는 '빅카인즈'를 통해 수집하였다. 빅카인즈는 중앙지, 경제지, 지역종합지, 방송사, 전문지 등 54개 언론사의 뉴스를 수집하여 제공하고 있다(빅카인즈). 가축 폐사 자료는 농업정책보험금융원의 가축재해보험 자료를 이용하였다.

뉴스 자료는 축산 부문의 피해를 보도한 모든 뉴스 중 중복 뉴스를 제거한 후, 단계별 전처리를 거쳐 활용하였다. 뉴스는 다음과 같은 과정으로 추출하였다. 1차적으로 '폭염' 또는 '더위', '고온'을 포함하는 뉴스를 검색하여 폭염과 관련된 모든 뉴스를 추출하였고, 2차적으로 '피해', '질환', '폐사', '죽음', '사망', '장애', '사고', '영향' 등의 단어를 포함하는 뉴스를 검색하여 피해와 관련된 뉴스를 추출하였다. 마지막으로 '젖소', '한우', '육우', '닭', '오리', '메추리', '양계', '육계', '돼지', '축산', '가금류' 등의 단어를 포함하는 뉴스를 추출하였다. 그 결과 3,192개의 뉴스를 추출할 수 있었다.

추출된 뉴스들 중 어부정 뉴스, 언론사가 제목이나 내용을 일부 바뀌가며 같은 내용을 반복해서 송고하는 뉴스들을 제거하기 위해 내용의 유사도를 분석하여 내용이 50% 이상 유사한 뉴스는 가장 먼저 등록한 뉴스를 제외하고 모두 제거하는 방법을 사용하였다. 3,192개의 뉴스 중 933개(약 29%)가 유사도 검사로 제거되었고, 2,259개의 뉴스가 연구에 사용되었다(표 1). 유사도 분석에는 R의 'quanteda' 패키지에 포함되어 있는 `textstat_simil` 함수를 이용하였다.

농업정책보험금융원의 가축재해보험 보장내용에 의하면, 가축재해보험은 축산농가를 위해 만들어진 보험상품으로 소, 말, 돼지, 가금류(닭, 오리, 꿩, 메추리, 타조, 거위, 칠면조, 관상조)의 피해를 보상하며, 돼지와 가금류에 대해서는 폭염추가특약을 통해 폭염에 의한 피해도 보상한다. 본 연구는 돼지와 가금류 중 닭, 오리, 메추리의 폭염 피해 보상 내역을 연구에 사용하였다. 감사원(2019)의 농어업재해보험 운영실태 공개문을 분석한 결

표 1. 연구에 사용된 뉴스 및 가축 폐사 자료 현황 및 출처

구분		2015년	2016년	2017년	자료 출처
뉴스(건수)		491	1,018	750	빅카인즈
폐사(마리)	돼지	4,819	38,652	73,835	농업정책보험금융원 가축재해보험
	닭	2,580,961	5,808,681	6,832,409	
	오리	81,977	196,777	282,333	
	메추리	12,023	86,713	71,537	
보험가입 마릿수 대비 폐사율(%)	돼지	0.05	0.40	0.71	저자 계산
	닭	1.31	2.55	2.83	
	오리	1.06	2.21	3.00	
	메추리	0.21	1.28	1.00	

표 2. 연도별 가축재해보험 보험가입 마릿수

(단위: 천마리)

축종	2015년	2016년	2017년
돼지	8,769	9,600	10,396
가금류	210,940	243,483	257,638
닭	197,365	227,813	241,057
오리	7,712	8,901	9,419
메추리	5,857	6,760	7,153
총 가축	219,900	253,400	268,400

* 한국농업재해보험협회(2016:205; 2018:51); 감사원(2019:12)을 토대로 저자 작성.

** 2016년과 2017년에서 음영 처리된 닭, 오리, 메추리 보험가입 마릿수는 2015년 가금류에서 각 축종이 차지하는 비율을 토대로 저자가 추정한 값임.

과, 닭, 오리, 메추리의 보험가입 마릿수는 가축재해보험에 가입된 전체 가금류의 99.9%를 차지하였다. 따라서 세 종의 자료는 가금류를 대표하는 자료로 간주할 수 있다.

가축 폐사 수(피해 보상 마릿수)는 매년 증가 경향을 보였다(표 1). 가축재해보험 보험가입 마릿수 역시 지속적인 증가를 보였다(표 2). 2015년 보험에 가입된 총 가축 수는 약 2억2천만 마리아고, 2017년에는 약 2억6천8백만 마리로 1.2배 증가하였다. 가축의 폐사 수는 보험가입 마릿수의 증가와 연관되어 있을 가능성이 있기 때문에 폐사 수는 매년의 축종별 보험가입 마릿수에 대한 비율로 변환하여 사용하였다. 2016년과 2017년의 축종별 보험가입 마릿수는 획득이 어려워 2015년 자료의 가금류와 각 축종별 비율을 이용하여 추정된 값을 사용하였다. 보험가입 마릿수를 고려했을 때도 축종별 폐사율은 메추리를 제외한 나머지 축종들이 모두 2017년에 폐사율이 가장 높게 나타났다(표 1).

본 연구는 돼지, 닭, 오리, 메추리에서 발생한 폐사 수

를 종합하여 축산 부문의 일별 피해 규모로 사용하였다. 하지만 돼지, 닭, 오리, 메추리는 군집의 크기에 큰 차이가 있기 때문에 폐사 수를 단순 합산하여 축산 부문의 피해로 간주하는 것은 불가능하다. 따라서 모든 축종을 고려한 일별 피해 규모는 축종별로 폐사 수를 정규화한 후, 정규화된 폐사 수를 평균하여 계산하였다. 축종별 폐사 수의 정규화에는 표준점수(Z-Score)를 사용하였으며 일별 피해 규모의 표준점수 계산에는 다음 식 (1)을 이용하였다.

$$z = \sum((x_i - \mu_i) / \sigma_i) / n \quad (1)$$

z는 주어진 일자에 대한 전체 축종의 폐사 규모이며, x_i 는 특정 축종(i)에 대한 일 폐사 수, μ_i 는 i 축종의 평균 폐사 수, σ_i 는 i 축종 폐사의 표준편차이다. n = 4로 전체 축종 수이다.

실제 가축 폐사와 뉴스에 언급된 피해 지역의 비교는

표 3. 시기별 연평균 폐사 수(2015~2017년)

(단위: 마리)

시기 구분	돼지	닭	오리	메추리
초기(5월 1일 ~ 7월 15일)	7,614	818,554	8,804	-
중기(7월 16일 ~ 8월 15일)	28,386	3,936,046	159,665	56,758
후기(8월 16일 ~ 9월 30일)	3,101	319,416	18,560	-

2017년을 대상으로 하였다. 뉴스에서 가축에 대한 피해가 언급된 지역들을 지도로 구축하였으며, 피해 지역은 시군과 시도 단위로 정리하였다. 뉴스에 나타난 피해 지역의 추출은 연구자가 뉴스의 내용을 읽고 정리하였다. 뉴스의 텍스트 마이닝을 통해 피해 지역을 추정하는 것도 가능하겠지만, 연구자가 직접 정리하는 과정을 통해 뉴스에서 피해 지역을 기술하는 방식을 이해하고 시사점을 도출하는 것이 관련 분야의 연구자들에게 더욱 유용한 정보를 제공할 수 있을 것으로 판단하였기 때문에 수작업을 통한 자료 수집 방법을 채택하였다. 가축 폐사 및 뉴스에서 언급된 피해 지역의 공간적 분포에 대한 이해를 돕기 위해 가금류와 돼지의 사육현황 정보를 활용하였으며, 이 정보는 농림축산식품부 농업경영체 등록 정보 조회 서비스(<https://edu.agrix.go.kr>)로부터 획득하였다.

2. 연구방법

가축 폐사와 축산 피해를 다룬 뉴스의 건수 증감 패턴 사이에 관계 분석을 위해 일별 폐사 수와 뉴스 건수의 상관관계를 분석하였다. 먼저 일 최고기온에 따른 뉴스 건수와 폐사 수의 변화를 비교하였다. 일 최고기온을 이용하여 분석한 이유는 가축 폐사가 기온과 연관되어 있기 때문에 실제 피해가 급격하게 증가하기 시작하는 시점과 피해가 정점에 이르는 시기를 파악하기 용이하기 때문이다. 이 시점들을 중심으로 뉴스 건수의 변화를 살펴보면 폐사 수와 뉴스의 증감 패턴 사이에 관계를 파악할 수 있을 것으로 기대하였다.

폭염 피해가 증가하는 시기와 뉴스가 증가하는 시기의 관계는 교차상관(cross correlation)을 이용하여 분석하였다. 교차상관 분석은 축산폐사와 뉴스 건수의 상관관계에 시차(time-lag)가 존재하는지 여부와 시차의 정량적 측정을 위해 채택하였다. 교차 분석에 앞서 Augmented Dickey-Fuller test와 Kwiatkowski-Phillips-Schmidt-Shin

test를 수행한 결과 두 자료 모두 정상성(stationarity)이 없고, 평균이 일정하지 않았기 때문에 차분을 통해 자료의 정상성을 확보한 후 교차분석을 수행하였다. 교차상관 분석에는 R의 'tseries' 패키지를 이용하였다.

기온과 시기에 따른 뉴스의 주요어 변화를 분석하고, 워드클라우드(wordcloud)로 시각화하였다. 기온별 분석은 기온 상승 및 폐사 증가에 따른 뉴스의 내용적 변화를 파악하기 위하여 수행하였으며, 시기별 분석은 폐사가 증가하는 시기와 감소하는 시기에 뉴스의 내용적 차이를 비교하기 위하여 수행하였다. 시기별 분석에서는 매년 전체 기간(5월 1일~9월 30일)과 여름 초기(5월 1일~7월 15일), 여름 중기(7월 16일~8월 15일), 여름 후기(8월 16일~9월 30일)에 대한 주요어의 빈도를 분석하고 워드클라우드를 시각화하였다. 여름 초기는 폐사가 발생되기 시작하는 시기이며, 여름 중기는 폐사가 증가하는 시기, 여름 후기는 폐사가 감소하는 시기로 볼 수 있다(표 3).

워드클라우드에는 축산 부문의 피해와 연관된 380개의 단어로 구성된 사전을 사용하였다. 단어 사전은 연구에 사용된 뉴스에서 상대적으로 높은 빈도를 보인 단어들 중 '폭염', '무더위', '더위', '가마솥더위', '열대야' 등 기상과 연관되거나 숫자, 지역, 농작물 등 축산 피해와 무관한 단어, 2017년의 '살충제 달걀'과 같이 특정한 사회적 이슈와 관련된 단어들을 제외하고 구축하였다. 워드클라우드에는 R의 'quanteda' 패키지를 이용하였다.

III. 연구 결과

1. 기온과 가축 폐사의 전반적 변화 경향

우리나라의 5~9월 일 최고기온의 전국 평균은 2015년에 상대적으로 낮았고, 2016년과 2017년에 높게 나타났다. 2016년의 중앙값은 2017년에 비해 0.4°C 낮았지만 최

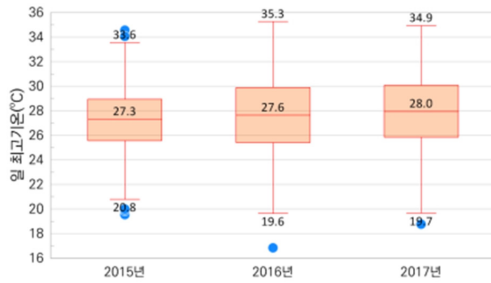


그림 1. 일 최고기온 전국 평균의 분포(5월~9월)

표 4. 일 최고기온의 전국 평균이 기준 기온을 넘는 일수

기준	2015년	2016년	2017년
30°C 이상	26	38	39
31°C 이상	16	34	27
32°C 이상	11	27	13
33°C 이상	6	20	6

* 기상자료개방포털 종관기상관측 자료를 활용하여 저자 작성.

표 5. 정규화 된 폐사 수(Z-Score)와 연도별 폐사 발생 일수

구분		2015년	2016년	2017년
Z-Score	전체	-0.16	0.04	0.12
폐사 발생 일수	돼지	24	62	76
	닭	32	61	59
	오리	17	42	32
	메추리	1	3	5
	전체	37	72	79

고기온은 35.3°C로 0.4°C 높았다(그림 1). 일 최고기온이 특정 기온을 넘어서 일수를 비교한 결과 30°C 이상은

2017년이 2016년 보다 하루 더 많았지만, 31°C 이상 일수 부터는 2016년이 가장 높게 나타났다. 33°C를 넘는 날을 비교하면 2015년과 2017년은 6일로 같았고, 2016년은 20일로 다른 해에 비해 2배 이상 많았다. 2016년이 다른 해에 비해 극한 기온이 더 자주 나타났음을 알 수 있다(표 4).

전체 가축에 대해 정규화 된 폐사 수는 모두 2017년에 가장 높게 나타났다. 정규화 된 폐사 수는 2015년에 -0.16으로 평균 보다 낮았고, 2016년에는 0.04, 2017년에는 0.12를 나타냈다(표 5). 극한 고온 현상이 나타난 일수가 2016년에 더욱 많았음에도 불구하고, 폐사 규모가 2017년에 가장 높게 나타난 것은 매년 극한 고온이 나타난 지역과 피해를 입은 농장의 규모에서 차이가 있기 때문일 수 있다. 이러한 가능성은 연도별 폐사 발생 일수의 비교를 통해서 확인할 수 있다.

연도별 폐사 발생 일수는 2015년에 37일, 2016년에 72일, 2017년에 79일이었다. 돼지와 메추리의 폐사 발생 일수는 2017년에 더욱 많았지만 닭과 오리의 폐사 발생 일수는 2016년에 각각 61일과 42일로 가장 많았다. 하지만 표 1에서 2016년 닭과 오리의 폐사율은 2017년에 비해 각각 0.28%p와 0.79%p 낮았다. 이는 매년의 폐사 규모가 고온이 나타난 지역과 실제 폐사 발생 농가의 규모에 의해 영향을 크게 받는다는 것을 의미한다. 아울러 보험가입 마릿수와 마찬가지로 농가 역시 점차 증가하였는데(감사원, 2019:12), 이를 감안하면 2017년에 폐사 발생 일수 증가는 보험가입 농가수의 영향을 받았을 가능성이 높다. 따라서 연도별 폐사 발생 일수의 단순 비교는 어려움이 있다.

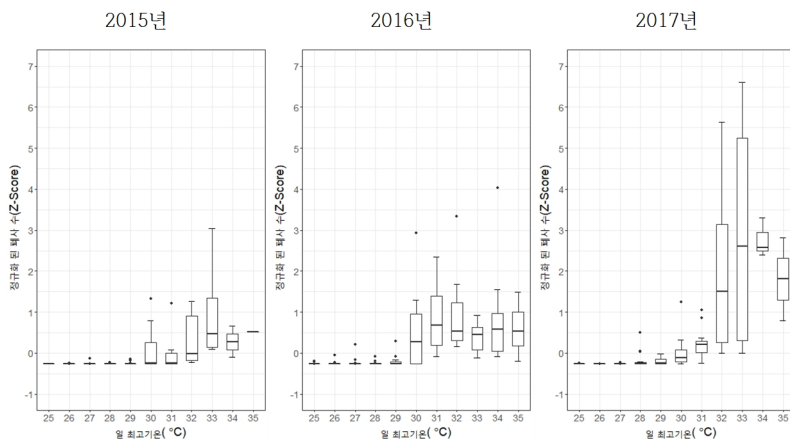


그림 2. 연도별 일 최고기온과 폐사와의 관계

표 6. 시기별 가축 전체의 정규화 된 폐사 수(Z-Score)

시기 구분	2015년	2016년	2017년
초기(5월 1일 ~ 7월 15일)	- 0.117	- 0.104	- 0.074
중기(7월 16일 ~ 8월 15일)	0.036	0.190	0.257
후기(8월 16일 ~ 9월 30일)	- 0.076	- 0.051	- 0.056

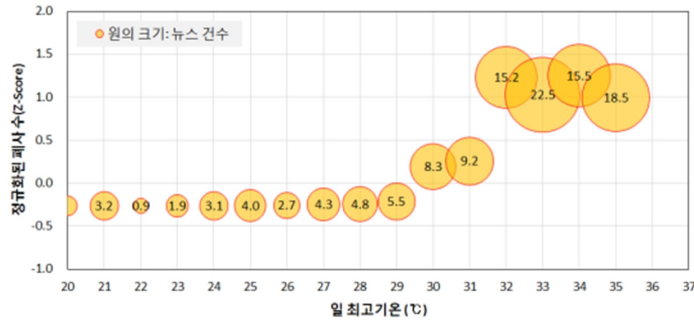


그림 3. 기온에 따른 일 평균 가축 폐사 수와 뉴스 건수(2015~2017년)

기온 1°C 구간 마다 발생한 폐사 수를 살펴보면, 매년 발생 규모 측면에서 차이가 나타나고 있으나, 대부분의 폐사는 매년 약 30°C 이상에서 발생하고 있으며, 이는 기온이 상승함에 따라 폐사 규모가 증가하는 경향이 있음을 의미한다(그림 2). 시기적으로는 7월 중순부터 8월 중순에 피해가 집중되고 있다. 매년 초기에는 가장 낮은 음의 Z-Score가 나타나고 중기에는 양의 Z-Score가 나타나며 후기에는 초기에 비해 높은 Z-Score가 나타난다. 그러나 후기에는 여전히 음의 값을 보인다(표 6).

분석 결과를 종합하면 가축 폐사 수와 폐사 일수는 매년 보험에 가입한 대상과 극한 고온의 공간적 분포 변화에 영향을 받기 때문에 연도별 피해량의 정량적 비교를 목적으로 한 연구에 활용하기에는 한계가 있다. 그러나 기온에 따른 폐사 발생이 매년 유사한 패턴을 보이기 때문에 고온에 의해 발생한 폐사와 뉴스 증감 패턴 사이에 관계를 분석하는 것은 가능한 것으로 판단된다. 또한 시기에 따른 폐사 발생 역시 그 규모에 차이는 있지만 동일한 증감 패턴을 보이고 있으므로 각 시기에 발표된 뉴스의 특성을 폐사 발생 특성과 비교하여 분석하는 것이 가능한 것으로 판단된다.

2 가축 폐사와 뉴스 증감 패턴의 관계

2015년부터 2017년까지 발생한 가축 폐사 규모를 기

온에 따라 정리하면 한국에서는 일 최고기온의 전국 평균이 30°C를 넘어설 때 가축 폐사가 증가하는 경향이 뚜렷하게 나타났다. 가축의 폭염 피해를 다루는 뉴스는 가축 폐사가 증가하는 기온에서 동일하게 증가하고 있었다. 하지만, 폐사 규모와 뉴스 건수의 정점에서는 차이가 있었다. 폐사 규모는 34°C에서 정점이 나타난 반면 뉴스는 33°C에서 정점이 나타났다(그림 3). 이 결과는 뉴스가 가축 폐사의 최고점을 전·후하여 크게 증가할 가능성을 보여준다.

교차상관 분석 결과에 의하면 뉴스는 가축 폐사보다 후행하여 증가하는 경향이 있었다. 그림 4는 가축 폐사와 뉴스의 교차상관 분석 결과이다. +6일에서 두 변수 사이에 양의 상관관계가 나타나고 있다. 그림 5는 시차에 따른 가축 폐사와 뉴스의 일변화를 보여준다. 두 자료를 당일(Lag-0, 그림 5의 (a))로 비교했을 때 뉴스의 정점들이 가축 폐사의 정점에 비해 뒤쪽에 분포하고 있는 것을 볼 수 있다. 가축 폐사 6일 후에 뉴스가 증가한다고 가정하여 6일의 지연 효과를 부여하면(Lag-6, 그림 5의 (b)) 당일 자료를 비교했을 때에 비해 정점이 유사해지는 것을 확인할 수 있다. 이런 현상은 2015년과 2016년에 가축 폐사가 급격하게 상승하는 초반에 두드러지게 나타난다. 2017년에는 가축 폐사가 급격하게 증가한 첫 시기와 뉴스의 최대 정점 사이에 지연 효과를 부여한 이후에도 약 6일 간의 추가적인 차이가 존재한다. 하지만

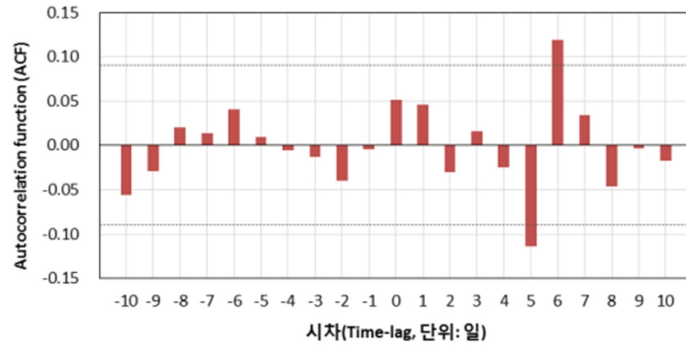


그림 4. 가축 폐사와 뉴스의 교차상관 분석 결과

* 점선은 유의수준을 의미함.

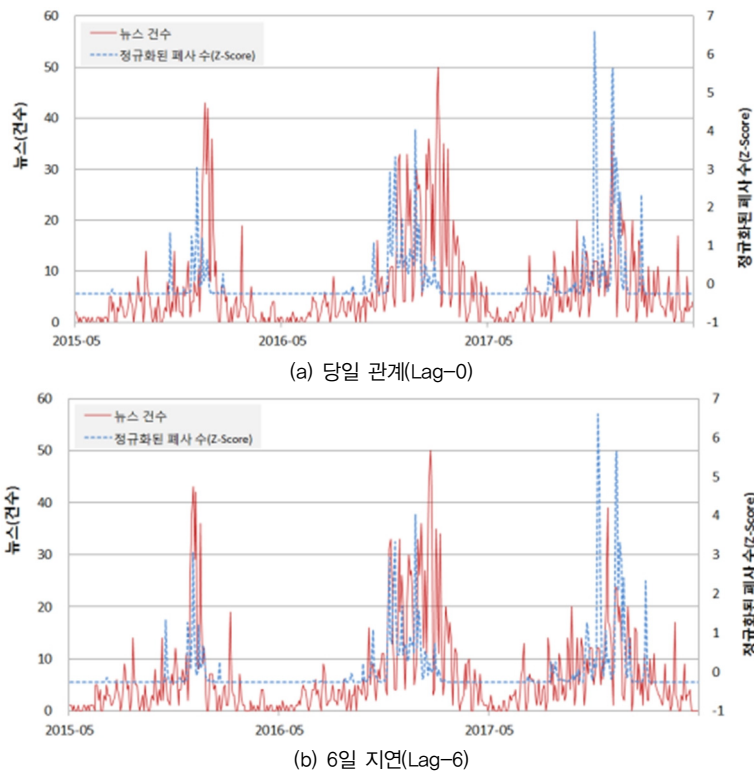


그림 5. 시차에 따른 가축 폐사와 뉴스의 일변화 비교

가축 폐사가 크게 증가한 두 번째 정점에서는 뉴스의 정점이 일치하는 경향이 있다. 이 결과들은 뉴스가 가장 많았던 기온이 가축 폐사가 최고를 보였던 기온에 비해 1°C 낮거나 높았던 이유가 가축 폐사가 발생한 후 이를 다루는 뉴스 기사들이 약 6일 정도 후에 나타났기 때문이라는 것을 보여준다.

3. 기온과 시기에 따른 뉴스 주요어의 변화

앞서 기온이 증가함에 따라 가축 폐사가 증가하고, 비슷한 시기에 축산 피해를 다루는 뉴스 역시 증가하는 것을 확인하였다. 이때 뉴스들이 다루는 주요 내용을 살펴보면 31°C 까지는 고온 피해로 인한 물가 불안정에 대한

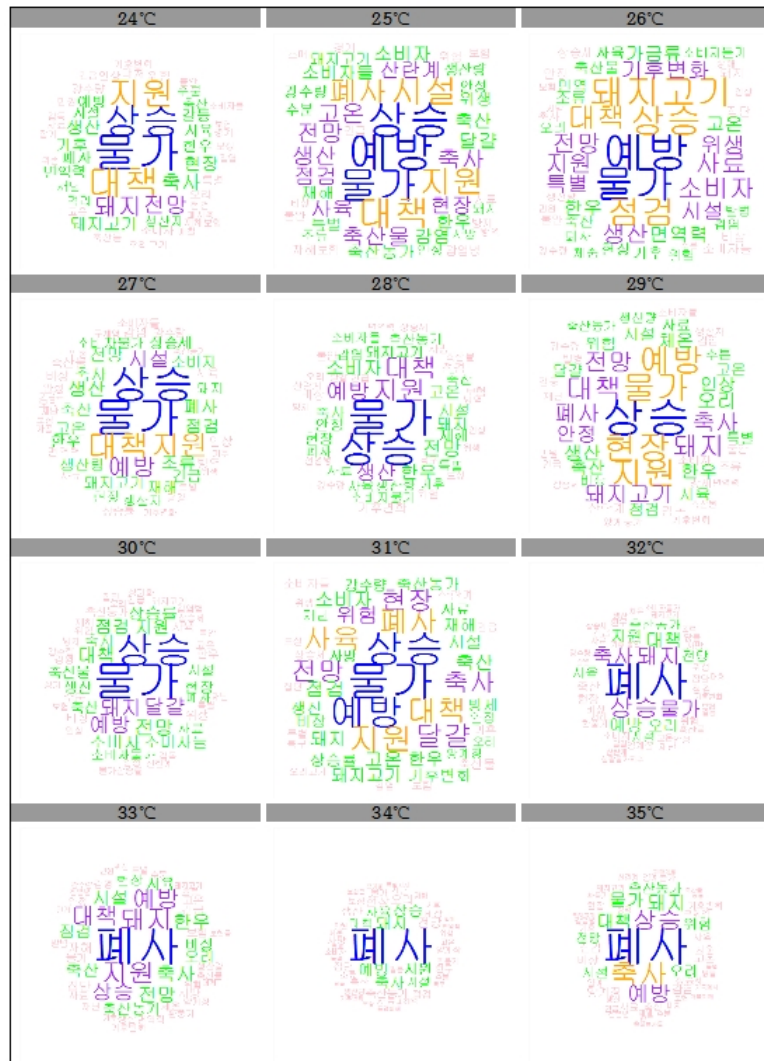


그림 6. 기온에 따른 뉴스 주요어의 변화

우려를 다루고 있으며 32°C 부터는 폐사와 관련된 내용이 주를 이루었다(그림 6). 가축 폐사가 최고조에 이르는 34°C의 키워드를 보면 다른 기온에 비해 폐사가 특히 강조되고 있다. 33°C와 35°C에서는 폐사가 여전히 가장 주요어지만 폐사 외에도 예방, 대책, 지원 등 다양한 주제들이 나타나고 있다. 33°C와 35°C의 결과는 가축 폐사가 정점에 이른 후 약 6일 후에 증가하는 기사들이 폐사와 연관된 다양한 주제들을 이용하여 뉴스를 생산하고 있다는 것을 보여준다.

그림 7은 시기에 따른 뉴스 주요어의 차이를 보여준

다. 2015년에는 초기에 ‘예방’, ‘지원’, ‘대책’ 등과 관련된 뉴스들이 주를 이루다가 중기가 되면 ‘폐사’와 ‘예방’을 다루는 기사들이 증가하였다. 후기에는 ‘사료’와 ‘상승’이 주요어를 이루고 있었다. ‘사료’는 사료섭취량 감소에 따른 생산성 저하 또는 폐사 등에 대한 우려와 연관되어 있었고 ‘상승’은 물가상승과 연관되어 있었다. 2016년의 시기별 주요어는 2015년과 대체로 유사한데 후기까지 ‘폐사’가 나타나고 있으며 물가상승과 관련된 주요어의 비중이 2015년에 비해 높은 것이 특징이다. 2017년의 뉴스에서는 모든 시기에 ‘물가’가 주요어로 선정되었다. 중

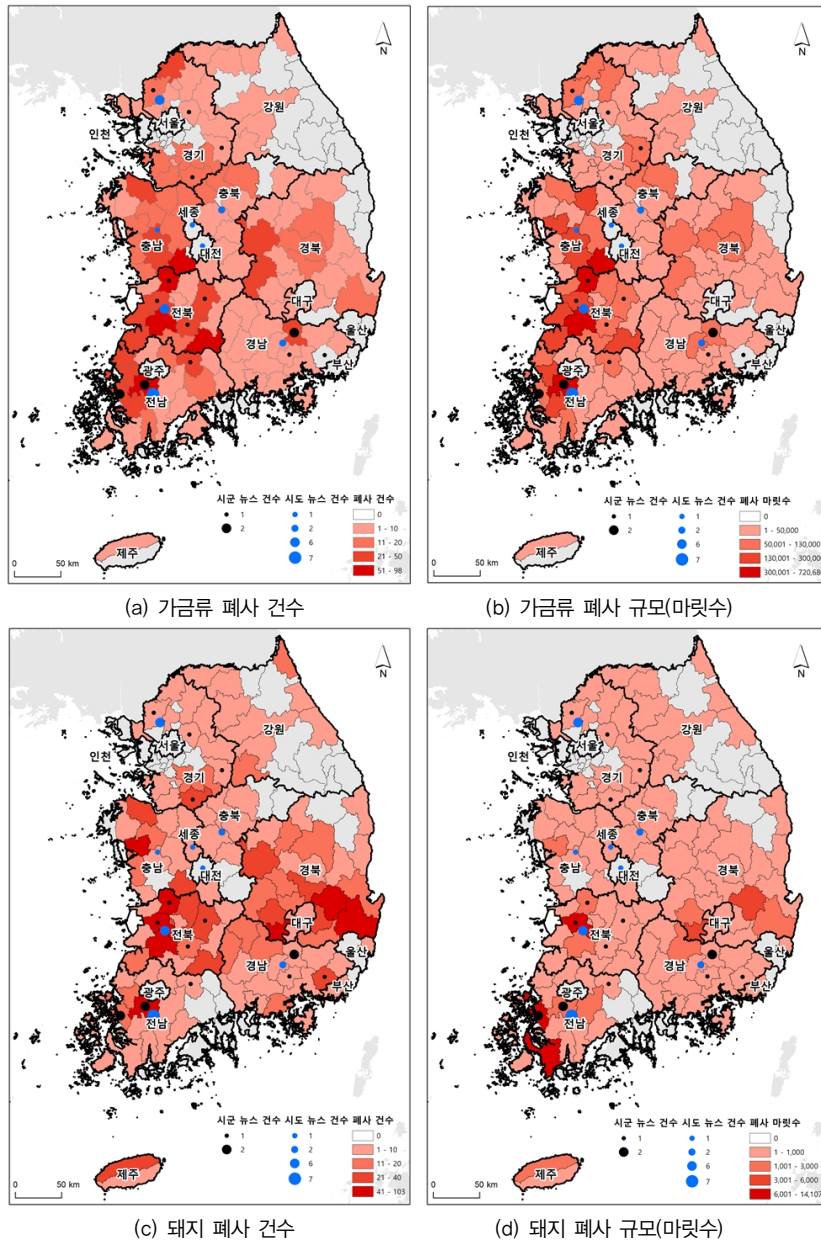


그림 8. 2017년 뉴스에 언급된 피해 지역과 실제 피해의 공간적 분포 비교

서는 폭염으로 인한 축산 농가의 피해를 다음과 같이 기술하고 있다.

“폭염에 가축 폐사가 속출하고 있다.

지난 4일 기준으로 모두 255농가에서 32만498마리의 닭, 오리, 돼지가 폐사했다. 가축별로 닭 29만6730마리, 오리

2만2872마리, 돼지 896마리다.

목포, 여수, 완도를 제외한 도내 19개 시·군에서 피해가 발생했다. 닭·오리 사육량이 많은 나주가 7만8223마리로 가장 많고 곡성 4만6901마리, 영암 4만4423마리, 무안 3만558마리 등이다.”

(광주일보, 2017년 8월 8일자, 3면)

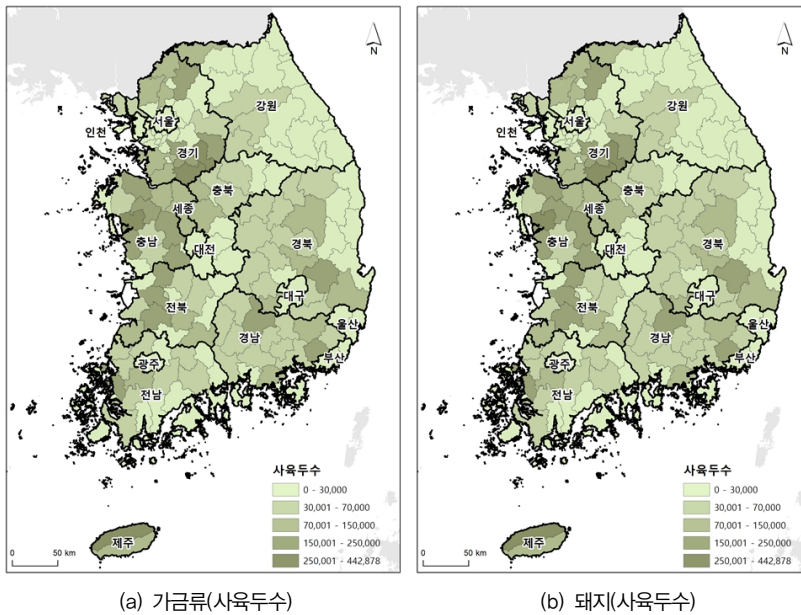


그림 9. 2017년 가금류와 돼지의 사육 두수 분포

피해가 발생한 19개 지역 중 피해가 컸던 나주, 곡성, 영암, 무안만을 언급하고 있다. 다른 15개 지역은 뉴스를 통한 파악이 어렵다. 반면 피해가 발생하지 않은 목포, 여수, 완도가 기사에 포함되어 있다. 이는 뉴스로부터 피해 지역을 유추하기 위한 텍스트 마이닝 과정에서 제외 지역이 포함되어 있을 수 있으므로 ‘제외’ 등의 명사를 고려하여 오분류에 대한 주의를 기울일 필요가 있다는 것을 시사한다.

IV. 토론

본 연구의 결과는 뉴스 빅데이터를 축산 부문의 사회적 이슈를 모니터링 하고자 할 때 관련 뉴스의 건수 보다는 주요어의 분석이 유용할 수 있다는 것을 보여준다. 특히 주요어 분석은 가축 폐사의 모니터링에 유용할 가능성이 높다. 뉴스 건수는 큰 규모의 폐사가 발생한 후 6일 후에 증가하는 경향이 있지만 ‘폐사’라는 단어가 뉴스의 주요어를 차지하는 시기는 실제 폐사가 집중되는 시기와 일치하기 때문이다. 주요어를 기반으로 할 때 예상되는 한계는 특정한 이슈가 사회적 관심사로 나타나게 될 때이다. 이 경우 그 이슈에 밀려 다른 현상들은

뉴스에서 잘 다루어지지 않기 때문이다.

7월 중순 이전의 뉴스들은 주로 예방, 지원, 대책 등으로 내용을 구성하고 있었다. 폐사가 집중되기 시작하는 7월 중순부터 8월 중순까지는 폐사가 주요어를 차지한다. 그리고 그 이후에는 축산 피해에 의한 물가 변화 기사들이 주를 이룬다. 이러한 시간적 패턴은 매우 일반적이지만 특정 사회 이슈가 있는 경우에는 패턴에 변화가 나타나기도 한다. 2017년은 달걀에서 살충제가 발견된 이슈로 인해 모든 시기에 물가가 가장 주요한 단어였다. 그림 10은 2018년의 시기에 따른 뉴스 주요어의 변화를 보여준다. 시기에 따른 주요어가 2015, 2016년과 같이 예방 - 폐사 - 물가로 구성되어 있는 것을 확인할 수 있다. 이를 통해 2017년의 사례가 특정 이슈에 의해 주요어의 패턴에 변화가 발생한 사례라는 것을 확인할 수 있다. 특정 이슈로 인해 다른 이슈에 대한 모니터링이 어렵다면, 그 해에는 특정 이슈와 연관된 단어를 제외하고 주요어를 분석하는 것도 특정 이슈의 영향을 최소화하는 한 방법이 될 수 있을 것이다.

본 연구는 전국 평균 기온을 사용하였기 때문에 연구 결과에서 제시된 기온이 가축 폐사가 발생하는 기온을 직접적으로 지시하지 않으며, 마찬가지로 뉴스가 증가하는 기온을 직접적으로 지시하지도 않는다. 폐사가 증



그림 10. 2018년 시기에 따른 뉴스 주요어의 변화

가하는 실제 기온은 각 지역의 주요 축종과 시설 특성 등에 따라 상이하며, 기사 역시 각 지역의 기온에 영향을 받을 수 있기 때문에, 이를 파악하기 위해서는 지역적인 규모에서의 기온과 가축 폐사, 뉴스의 관계를 분석할 필요가 있다. 본 연구에서 제시된 기온에 따른 분석 결과는 기온 상승에 대한 전반적인 반응 측면에서 이해할 필요가 있다.

V. 결론

본 연구에서는 기온 상승에 의해 가축이 폐사하면, 축산 피해를 다루는 뉴스 역시 증가하는 패턴이 있다는 것을 확인하였다. 가축 폐사가 집중되는 기온 구간(일 최고기온의 전국 평균 기준 32~35°C)에서 축산 피해를 다루는 뉴스는 이전 기간에 비해 두 배 이상 증가하였다. 하지만 뉴스의 건수가 가장 많이 증가하는 시기는 가축이 폐사한 후 6일 후에 나타나는 경향이 있었다.

가축 폐사가 증가하는 기온 구간에서 뉴스의 가장 주요어는 폐사였다. 많은 폐사가 발생한 후 약 6일의 시간이 경과하면 뉴스는 폐사 외에도 예방, 대책, 지원 등 다양한 주제들을 다루고 있었지만, 여전히 가장 주요하게 다룬 주제는 폐사였다.

뉴스 빅데이터를 이용하여 가축 피해, 특히 폐사에 대한 모니터링 방안을 설계한다면 뉴스 건수와 주요어의 특성을 고려했을 때 주요어 분석이 유용한 방안이 될 것으로 판단된다. 뉴스 건수는 폐사 이후 일정한 지연시간을 가진 후에 증가하지만 ‘폐사’라는 주요어는 실제 가축 폐사와 동일 시점에 많이 다루어지기 때문이다. 다만 주요어의 경우 특정한 사회적 이슈의 등장에 의해 영향을 받기 때문에 이에 대한 충분한 고려를 필요로 한다.

뉴스에서 피해 지역에 대한 정보를 추출하고 실제 피해 지역과의 공간적 분포 차이를 비교한 결과는 뉴스를 이용한 공간적 연구의 가능성과 극복해야 할 과제를 보여주었다. 많은 피해가 집중된 지역들을 중심으로 뉴스에서 언급하고 있다는 것을 확인한 것은 뉴스를 이용한 공간적 연구의 가능성을 보여준 것이라고 판단된다. 그러나 뉴스에서 피해가 발생하지 않은 지역을 함께 언급하는 사례가 있고, 많은 피해 지역들 중 대표적인 몇몇 지역만을 언급한다는 것은 뉴스를 이용한 공간적 분포 연구에서 반드시 고려해야 할 한계이다. 다만 본 연구에서는 2017년 여름 사례만을 대상으로 분석을 수행하였기 때문에 뉴스를 이용한 공간적 연구의 한계를 충분히 확인하지 못하였다. 따라서 이에 대해서는 추가적인 연구를 통해 자료의 특성을 보다 면밀하게 파악할 필요가 있다.

본 연구의 결과는 뉴스에서 다루어지는 축산 피해와 실제 피해의 관계에 대하여 분석하였다. 축산 피해가 폐사로 국한되기는 하였지만 뉴스 빅데이터를 활용하여 축산 부문의 피해를 모니터링하기 위한 기초 정보를 제공하고 있다. ‘폐사’는 피해가 집중되는 시기를 모니터링하는 데 유용한 키워드로 활용할 수 있으며, ‘예방’과 ‘물가상승’은 축산 피해가 일어나기 전과 소강상태로 접어드는 시기를 탐지하는 데 유용한 키워드로 활용할 수 있다. 실제 피해와 뉴스 건수 사이에 지연시간이 존재한다는 점은 뉴스를 이용한 모니터링 과정에서 뉴스 건수의 정점보다는 증감 패턴을 주요 모니터링 요소로 활용해야 한다는 점을 시사한다. 시기에 따른 주요어의 변화, 그리고 사회적 이슈에 의한 주요어의 교란 가능성은 뉴스의 활용 과정에서 지속적인 키워드 사전의 갱신과 관리가 필요하다는 것을 시사한다.

참고문헌

- 갑사원 2019, 「농어업재해보험 운영실태 공개문」, 서울: 갑사원.
- 국립재난안전연구원 2017, 「비정형 빅데이터 재난안전 정보 패턴 분석」, 울산: 국립재난안전연구원.
- 국립축산과학원 2016, 「폭염대비 고온기 가축 및 축사 관리 요령」, 수원: 국립축산과학원.
- 김동우·박종철·장동호, 2017, “표준강수지수와 정규식생지수를 활용한 봄 가뭄 탐지 가능성 분석” 한국지리학회지, 6(2), 165-174.
- 노병준·서정순·이종욱·정용화·박대희, 2016, “키워드 네트워크를 통한 구제역 동향 분석” 한국인터넷정보학회 학술대회발표집, 17(1), 217-218.
- 농촌진흥청, 2018, 「폭염 장기화에 따른 단계별 예상피해 및 대책」, 수원: 농촌진흥청.
- 박준범·남궁미, 2019, “고령 보행자 교통사고에 영향을 미치는 환경요인에 관한 연구: 부산광역시를 중심으로” 한국지리학회지, 8(2), 289-303.
- 알 마룬·장동호·박종철, 2017, “산사태 분포 예측을 위한 로지스틱, 베이즈, Maxent의 비교” 한국지형학회지, 24(2), 91-101.
- 오성렬·이기하, 2014, “공간분포형 습윤지수를 이용한 유역 규모의 사면안정해석” 한국지리학회지, 3(2), 111-126.
- 이궁희, 2017, “뉴스 빅데이터를 이용한 경기 판단: 빅카인즈 뉴스 경기지수의 개발” 통계연구, 22(2), 67-94.
- 임현택·배상영·심규성·김정수·박성용·김용성, 2016, “재해 지도 활용성 증대를 위한 빅데이터 구축 및 적용 방안” 한국방재학회 논문집, 16(2), 205-2013.
- 장동호, 2007, “산사태 재해에 대한 삼림지 위험평가 분석모델 평가” 한국지형학회지, 14(4), 57-67.
- 장은미, 2006, “지리정보기반의 재해 관리시스템 구축(I) -민간 보험사의 사례, 태풍의 경우” 대한지리학회지, 41(1), 106-120.
- 정우석·오승희·이용태, 2019, “재난 빅데이터를 활용한 대형 복합재난 피해확산예측 시스템 구현에 관한 연구” 한국통신학회 논문지, 44(7), 433-442.
- 진대용·강성원·최희선·한국진·김도연, 2018, 「기후환경 이슈 분석을 위한 텍스트 마이닝 활용방안 연구」, 세종: 한국환경정책·평가연구원.
- 채여라·박종철·이주형, 2018, “수요자 중심 맞춤형 폭염 대응방안 마련” KEI 포커스, 6(9), 3-27.
- 최선화, 2014, “빅데이터 기반 재난 전조감지 모델” 한국방재학회 논문집, 14(2), 149-158.
- 하태환·권경석·이인복·김락우·여옥현·이상연·최희철·김종복·이준엽·전중환·우샘이·양가영, 2018, “하절기 자돈 고온 스트레스 평가를 위한 THI 지수 모델” 한국농공학회 논문집, 60(4), 113-122.
- 한국기후변화대응전략연구소, 2013, 「이상기온에 의한 재난 대응체계 구축방안」, 서울: 국립재난안전연구원.
- 한국농업재해보험협회, 2016, 「2016년 농업재해보험연간, 천안: 한국농업재해보험협회.
- 한국농업재해보험협회, 2018, 「2018년 농업재해보험 계간지 보호」, 천안: 한국농업재해보험협회.
- Ihm, H., Jang, K., Lee, K., Jang, G., Seo, M.G., Han, K., and Myaneng, S.H., 2017, Multi-source food hazard event extraction for public health, *Proceedings of 2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, February 13-16, Jeju, South Korea, 414-417.
- Rosenfeld, C.L., 1998, Recent Contributions by geographers to natural hazards mapping and mitigation, *Journal of the Korean Geographical Society*, 33(special issue), 721-728.
- Tarasconi, F., Farina, M., Mazzei, A., and Bosca, A., 2017, The role of unstructured data in real-time disaster-related social media monitoring, *Proceedings of 2017 IEEE International Conference on Big Data (Big Data)*, December 11-14, Boston, MA, USA, 3769-3778.
- 광주일보, 2017년 8월 8일자, “애플망고파파야커피 재배 ... 농사지도 새롭게 그린다”
- 기상자료개방포털, <https://data.kma.go.kr>
- 농림축산식품부 농업경영체 등록정보 조회 서비스, <https://edu.agrix.go.kr>
- 농업정책보험금융원 <https://www.apfs.kr>
- 빅카인즈, <https://www.bigkinds.or.kr>
- 교신 : 채여라, 30147, 세종특별자치시 시청대로 370 세종국책연구단지 B동, 한국환경정책·평가연구원 (이메일: yrchae@kei.re.kr)

Correspondence : Yeora Chae, 30147, Bldg. B, 370
Sicheong-daero, Sejong Special Self-Governing
City, Korea, Korea Environment Institute (Email:
yrchae@kei.re.kr)

투 고 일: 2019년 9월 2일

심사완료일: 2019년 11월 21일

투고확정일: 2019년 11월 21일