

빅데이터를 활용한 HPAI Virus 확산 예방 및 추적*

Prevent and Track the Spread of Highly Pathogenic Avian Influenza Virus using Big Data

최대우** · 이원빈***† · 송유한*** · 강태훈*** · 한예지***

한국외국어대학교 대학원 통계학과

요약

이 연구는 2018년도 정부(농림축산식품부)의 재원으로 농림식품기술기획평가원의 지원을 받아 수행된 연구¹⁾이다. 고병원성 조류인플루엔자(Highly Pathogenic Avian Influenza, HPAI)는 해외로부터 철새를 통해 유입되고 있으나 어떤 경로를 통해 농가에 확산하는지 정확히 밝혀진 바 없다. 그리고 발생 농가에서 유입되는 농가 간의 전이도 차량이 주요 원인이라고 추정할 뿐, 전파 주요 원인이 정확히 밝혀진 것은 아니다. 하지만 가장 빈번하게 농가에 방문하는 차량의 방문유형이 가축 운반 및 사료 운반과 같은 농가와 시설 간의 방문이기 때문에 발생 농가에 들른 차량과 시설의 관계를 분석할 필요가 있다.

본 논문 연구에서는 농림축산검역본부에서 제공하는 KAHIS(Korea Animal Health Integrated System) 데이터를 기반으로, HPAI Virus 전이의 주요 원인을 확인하고자 한다.

■ 중심어 : Sankey diagram, Doc2Vec, TF-IDF, KAHIS(국가가축방역시스템), HPAI(고위험성 조류인플루엔자)

Abstract

This study was conducted with funding from the government (Ministry of Agriculture, Food and Rural Affairs) in 2018 with support from the Agricultural, Food, and Rural Affairs Agency, 318069-03-HD040, and is based on artificial intelligence-based HPAI spread analysis and patterning. Highly Pathogenic Avian Influenza (HPAI) is coming from abroad through migratory birds, but it is not clear exactly how it spreads to farms. In addition, it is assumed that the main cause of the spread is the vehicle, but the main cause of the spread is not exactly known. However, it is necessary to analyze the relationship between the vehicles and the facilities at the farms where they occur, as the type of vehicles that visit the farms most frequently is between farms and facilities, such as livestock transportation and feed transportation.

In this paper, based on the Korea Animal Health Integrated System (KAHIS) data provided by Animal and Plant Quarantine Agency, the main cause of HPAI virus transfer is to be confirmed between vehicles and facilities.

■ Keyword : Sankey diagram, Doc2Vec, TF-IDF, KAHIS(Korea Animal Health Integrated System), HPAI(Highly Pathogenic Avian Influenza)

2020년 11월 16일 접수; 2020년 12월 16일 수정본 접수; 2020년 12월 23일 게재 확정.

* 이 연구는 2018년도 정부(농림축산식품부)의 재원으로 농림식품기술기획평가원 지원을 받아 수행된 연구임.

** 한국외국어대학교 자연과학대학 통계학과 교수

*** 한국외국어대학교 대학원 통계학과 석사과정

† 교신저자 (wonbeeny@gmail.com)

1) 318069-03-2HD040, 인공지능기반 HPAI 확산 원인분석 및 패턴화 연구*

I. 서론

고병원성 조류인플루엔자는 매년 지속해서 발생하고 있다. 연중 고병원성 조류인플루엔자 발생으로 인한 경제적 피해는 축산물 생산액의 약 20%로 추정된다. 2000년 이후 누적 손실액은 약 6천4백억 원에 달한다. 이러한 경제적 손실을 주는 고병원성 조류인플루엔자의 최초 유입은 철새를 통해 유입되고 있으나 그 이후 어떤 경로를 통해 농가에 감염을 시키는지는 정확히 밝혀지지 않았다. 감염된 철새의 분변 1g에는 십만에서 백만 마리의 가금류를 감염시킬 수 있는 고농도의 바이러스가 들어있다. 이러한 분변은 주로 오염된 차량이나 사람 등을 통해 전염이 일어난다고 알려졌다.

고병원성 조류인플루엔자 바이러스 전파는 1차 확산과 2차 확산이 있다. 바이러스의 1차 확산은 철새 분변으로 인해 야생조수, 차량, 인력, 농자재에 의해 농가에 전파되는 것으로 알려졌으나 명확한 근거는 미흡하다. 2차 확산 또한 농가에 증폭된 바이러스가 차량 및 인력에 의해 농가 간 전파되는 것으로 알려졌으나 이 또한 명확한 근거는 미흡하다.

농가를 방문하는 차량의 23%만이 축산차량이고, 나머지 70% 이상은 일반 차량 및 인구로 분석되었다. 하지만 축산차량의 경우 매년 주기적으로 농가에 방문하기 때문에 축산차량과 고병원성 조류인플루엔자와의 상관관계를 분석할 필요가 있다. 증폭된 바이러스는 농가에 주기적으로 방문하는 가축 운반 차량, 사료 운반 차량과 같은 차량에 의한 전파일 가능성이 크다.

본 논문 연구에서는 바이러스의 전파 원인을 찾기 위해 바이러스의 전파가 차량 및 인력의 요인이 크다는 가정으로 시작한다. 축산차량의 이동과 역학분석을 통해 바이러스의 전파 원인을 찾고자 한다.

본 논문 연구에서는 농림축산검역본부가 제

공하는 KAHIS(Korea Animal Health Integrated System) 데이터를 이용하였다. KAHIS 데이터는 농가에 방문하는 가축 운반 차량, 사료 운반 차량과 같은 축산차량의 이동 정보를 모은 데이터이다. 고병원성 조류인플루엔자 Virus에 감염된 농가와 연관성이 높은 시설 및 해당 시설의 방문유형을 분석한 뒤, 고병원성 조류인플루엔자를 전파할 것으로 예상하는 차량을 확인하였다. 이를 통해 해당 차량의 방역을 강화하여 고병원성 조류인플루엔자 바이러스의 확산을 줄일 수 있을 것이다.

II. 본론

2.1 KAHIS 데이터

2.1.1 KAHIS 데이터에 대해

KAHIS 데이터는 농림축산검역본부에서 운영하는 축산차량등록제에 의해 축산차량을 등록하고 GPS 단말기를 장착하여 차량의 출입정보를 수집, 관리하는 Sequence to Sequence 데이터이다.

농림축산검역본부에서 제공한 KAHIS 데이터의 구성은 축산차량에 등록된 차량을 기준으로 축산차량 1년 동안의 과거 경로들로 이루어져 있다. 본 연구에 사용된 데이터는 보안 때문에 시설 간의 이동을 제외한 경로를 삭제하였다. 즉, 시설이 아닌 장소는 포함되지 않은 데이터이다.

해당 데이터의 칼럼 정보는 다음과 같다.

- 방문일자 : 시설을 방문한 시간
- 시설구분 : 시설의 종류 (농가, 사료공장 등)
- 방문유형 : 시설을 방문한 이유
- 차량번호 : 축산차량에 등록된 차량
- 주소 : 시설의 주소

본 데이터는 축산차량에 등록된 차량을 기준으로 추출한 데이터이기 때문에 각 차량의 방문 유형은 차량별로 동일하다. A 차량은 가축 운반을 목적으로 a 농가, b 농가, c 농가를 방문하고, B 차량은 사료 운반을 목적으로 각 농가를 방문한다. 방문유형이 같을지라도 시설구분은 다를 수 있다. 방문유형은 차량이 기준이 되고 시설구분은 주소가 기준이 되기 때문이다. 같은 주소라면 시설구분은 항상 같다. 마지막으로 방문 일자는 최초 GPS가 기록된 시간을 기준으로 한다. 즉, 농가로 들어간 시간과 나간 시간의 정보는 알 수 없다.

2.2 데이터 전처리

2.2.1 차량별 이동 경로 추출

차량별로 Sequence한 이동 경로를 추출하였다. 이동 경로는 중간 경로가 생략된 상태로 시설 간의 이동만 알 수 있다. 예를 들어 충청남도 천안시에서 경기도 평택시로, 경기도 평택시에서 전라북도 정읍시로 이동하는 것과 같이 출발 지점과 도착 지점만 존재하는 형태이다. 차량별 Sequence한 이동 경로가 아닌 출발 지점과 도착 지점만 나열된 형태이다.

2.2.2 차량별 이동 경로 분할

차량별 출발 지점과 도착 지점이 Sequence하게 나열된 형태로 데이터를 전처리해준 뒤, 차량별로 2개의 지역을 나뉜다. 예를 들어 A 차량의 이동 경로가 a → b → c → d 라면 a → b로, b → c로, c → d로 경로를 나눈다. 이렇게 나누는 이유는 HPAI 바이러스에 감염된 농가에서 다음으로 이동한 시설들의 빈도수를 구하기 위해서이다. 빈도수가 큰 시설일수록 HPAI 바이러스에 감염된 농가와 연관성이 높을 것으로 예상된다. HPAI 바이러스의 확산을 줄이기 위해서는 해당 지역에 방역을 강화할 필요성이 있다.

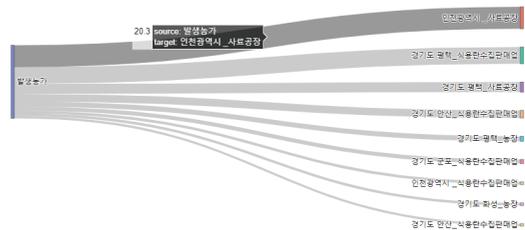
차량별 이동 경로 분할을 통해 만들 수 있는 데이터 테이블은 HPAI 바이러스에 감염된 농가와 방문유형의 관계에 대한 데이터 테이블을 만들 수 있다. 방문유형 테이블을 이용하면 HPAI 바이러스에 감염된 농가의 특징에 대해 알 수 있다.

차량별로 2개의 지역을 나누는 이유는 중간 경로가 생략됐기 때문이다. 3개 이상의 지역을 나눈다면 중간 경로의 생략으로 데이터의 신뢰성이 떨어질 것으로 예상된다.

2.3 Sankey diagram을 활용한 시각화

2.3.1 시설별 빈도수

Sankey diagram이란 Flow diagram의 한 종류로서 어떤 프로세스에서 주된 이동이나 흐름을 시각적으로 강조하며 어떤 항목의 기여도가 가장 높은지 찾는 데에 효과적인 시각화 방법이다.

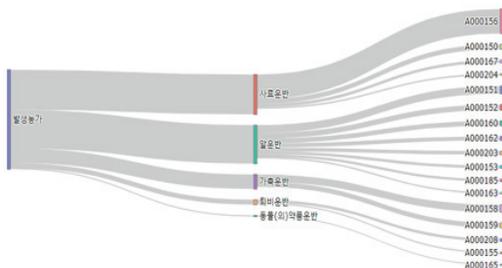


〈그림 1〉 시설별 빈도수 시각화

<그림 1>은 HPAI 바이러스에 감염된 농가에 방문한 차량이 이동한 시설별 빈도수의 관계를 보기 위해 전처리한 데이터를 Sankey diagram을 이용해 시각화한 그림이다. 왼쪽의 발생 농가는 HPAI 바이러스에 감염된 농가를 뜻하고, 오른쪽으로 뻗어 나가는 곳에 있는 인천광역시_사료공장, 경기도 평택_식용란수집판매업 등은 발생 농가에서 이동한 시설을 뜻한다. 전체 주소 정보는 데이터 보안상 생략하였다. Sankey diagram을 통해 알 수 있듯이 발생 농가에서 가장 빈번히 이동하는 시설은 인천광역시에 있는 사

료공장이고 다음으로는 경기도 평택에 있는 식용란수집판매업이다. 해당 농가는 주로 인천광역시와 경기도 평택에 있는 사료공장과 식용란수집판매업과 거래를 하고 있다.

2.3.2 방문유형별 빈도수



<그림 2> 방문유형별 빈도수 시각화

<그림 2>는 HPAI 바이러스에 감염된 농가에 방문한 차량의 방문유형별 빈도수와 방문유형별 차량의 빈도수를 Sankey diagram을 이용해 시각화한 그림이다. 왼쪽의 발생 농가는 HPAI 바이러스에 감염된 농가를 뜻하고, 오른쪽으로 뻗어 나가는 곳에 있는 사료 운반, 알 운반 등은 방문유형을, 맨 오른쪽에 있는 A000156과 같은 항목은 방문유형별 차량번호를 뜻한다. 차량번호는 데이터 보안상 생략하였다. Sankey diagram을 통해 알 수 있듯이 발생 농가에 가장 빈번히 방문하는 유형은 사료 운반 > 알 운반 > 가축 운반 순이고, 각각의 방문유형별 발생 농가에 가장 빈번하게 방문한 차량은 A000156, A000151, A000158 차량이다.

2.4 특정 지역과 연관된 차량 찾기

2.4.1 연구 방법론

Sankey diagram을 이용한 시각화는 발생 농가를 기준으로 시설구분 및 방문유형별 연관성을 분석하는 방법이다. 만약 발생 농가 외 여러 지역에서 HPAI 바이러스가 발견되었다고 한다면

여러 지역에 대한 데이터 테이블을 만들고 시각화를 해야 한다. 이러한 단점을 보완하기 위해 특정 지역과 연관된 차량을 찾는 3가지 방법론을 제시한다. 가독성을 위해 HPAI 바이러스에 감염된 지역을 위험 지역이라 하겠다.

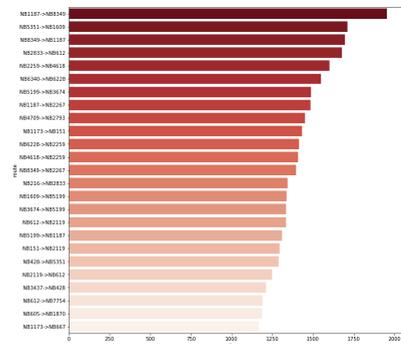
첫 번째 방법은 방문 비율을 구하는 알고리즘을 이용하여 위험 지역으로 예상되는 지점과 연관성이 높은 차량을 추출하는 방법이다.

두 번째 방법은 TF-IDF를 이용하여 위험 지역으로 예상되는 지점과 연관성이 높은 차량을 추출하는 방법이다.

세 번째 방법은 Doc2Vec 모델을 이용하여 위험 지역과 차량의 이동 경로 유사도를 구하는 방법이다.

특정 지역과 연관된 차량을 찾기 위해 세 가지 방법 중 두 가지 이상의 방법에 해당하는 차량을 위험 지역과 연관성이 높은 차량으로 지정하였다.

2.4.2 방문 비율 알고리즘



<그림 3> 방문 비율 결과

<그림 3>은 방문 비율 알고리즘의 결과이다. 방문 비율을 구하는 방법은 차량별 위험 지역에 방문한 수를 해당 차량이 들른 지역의 수로 나눈 값으로 정의한다. 예를 들어 위험 지역이 a, c, f라고 알려졌을 때, A 차량의 이동 경로가 c → b → f → a → g 라고 한다면 분자에 해당하

는 위험 지역에 방문한 수는 3이 되고, 분모에 해당하는 해당 차량이 들른 지역의 수는 5가 되어 방문 비율 값은 0.6의 값이 된다. 차량별로 방문 비율의 값을 계산하여 위험 지역과 연관성이 높은 차량을 추출하였다.

2.4.3 TF-IDF

TF-IDF는 정보 검색과 텍스트 마이닝에서 이용하는 가중치로, 여러 문서로 이루어진 문서 군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한지 나타내는 통계적 수치이다. 본 연구에서는 각 지역을 단어로 인식하고 차량이 이동한 경로를 문서로 인식하여 TF-IDF를 적용하였다. 식 (1)을 통해 알 수 있듯이 특정한 단어가 문서 내에서 단어 빈도가 높을수록, 그리고 전체 문서 중 그 단어를 포함한 문서가 적을수록 TF-IDF 값이 커진다. 식 (1)에서 x는 단어의 빈도수를, y는 문서를 뜻한다.

$$W_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right) \tag{1}$$

TF-IDF가 방문 비율을 계산하는 방법과 다른 점은 IDF를 통해 나타난다. IDF에 의해 모든 문서에 흔하게 나타나는 단어를 걸러내는 효과를 얻을 수 있기 때문이다.

2.4.4 Doc2Vec

Doc2Vec은 자연어처리에 사용되는 기술로 본 연구에서는 각 지역을 단어로 인식하고 차량의 이동 경로를 문장으로 인식하여 차량의 이동 경로와 위험 지역과의 유사도를 측정하기 위해 D2V 모델을 적용하였다.

〈표 1〉 D2V cosine similarity

Sent 1	Sent 2	similarity
아이유가 부른 요즘 노래 알아?	아이유의 신작 노래가 뭐야?	0.8975

〈표 1〉은 D2V 모델을 이용하여 문장 유사도를 구한 결과이다. 두 문장에 사용된 단어는 다르지만, 의미는 같으므로 약 0.9의 문장 유사도를 보인다. D2V 모델을 사용하면 단어의 순서와 의미를 동시에 고려하기 때문에 높은 유사도 결과를 보인 것이다. D2V 모델을 본 연구에 적용한 이유는 다음과 같다. 만약 어떤 차량에서 HPAI 바이러스가 검출된다면 해당 차량의 이동 경로와 연관성이 높은 차량을 찾는데 D2V 모델이 사용될 수 있다.

방문 비율 알고리즘과 TF-IDF는 위험 지역을 방문했는지 그렇지 않은 지만을 고려한 방법이다. 이동 경로 또한 중요한 요소이기 때문에 D2V 모델이 방문 비율 알고리즘과 TF-IDF의 단점을 보완하였다.

2.4.5 차량 추출

〈표 2〉 연관성이 높은 차량

	세 가지 방법론 적용	
	차량번호	Count
1	A000015	3
2	A000018	3
3	A000020	3
4	A000014	3
5	A000227	3
6	A000061	2
7	A000046	2
8	A000068	2

〈표 2〉는 세 가지 방법 중 두 가지 이상의 방법에 해당하는 차량을 추출한 결과이다. A000015, A000018, A000020 등의 차량은 세 가지 방법에 모두 확인되었다. 위험 지역과 연관성이 높은 차량이므로 해당 차량의 이동 경로를 분석하고, 해당 차량의 방역을 강화하고 해당 차량이 자주 방문한 시설에 방역을 강화할 필요가 있다.

본 결과는 실제 위험 지역으로 선정된 n개의 지역이 아닌 임의로 선정된 n개의 지역을 선정한 결과이다.

III. 결 론

3.1 연구 적용

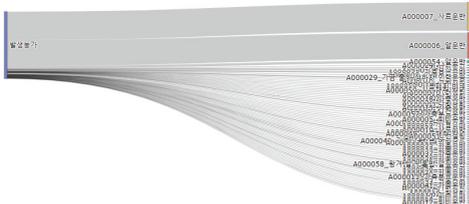
3.1.1 성능 평가의 한계

HPAI는 2018년 이후 발생하지 않았다. 그러므로 본 논문에서 제시한 연구 방법의 성능을 측정할 방법은 한정적이다. 또한, HPAI 바이러스에 감염된 농가에 최초로 HPAI 바이러스가 유입된 원인 또한 확실하지 않기 때문에 Sankey diagram을 이용한 발생 농가와 시설 및 방문유형의 빈도수와 세 가지 방법론을 이용한 HPAI와 연관성이 높은 차량 찾기의 성능을 평가할 지표는 없다.

본 연구에서는 HPAI 바이러스에 감염된 농가에 가장 많이 방문한 차량이 HPAI 바이러스 유입의 가장 큰 원인이라는 가정을 추가하였다. 이 가정을 세 가지 방법론에 따라 HPAI와 연관성이 높은 차량을 검출하는 방법이 유효한지를 판단하는 기준으로 정하였다.

3.1.2 데이터 적용 결과

결과 비교를 위해 세 가지 방법론과 Sankey diagram 결과를 적용하기 위해 경기도 화성에 있는 박○준 농가에 방문한 데이터로 세 가지 방법론에 적용해보았다. 가독성을 위해 박○준 농가를 발생 농가로 정의하겠다.



〈그림 4〉 차량 빈도수 Sankey diagram

〈그림 4〉는 발생 농가에 방문한 차량의 빈도수를 Sankey diagram으로 시각화한 그림이다. 사료 운반을 목적으로 A000007 차량이 가장 많이 방문했고 월평균 방문 빈도는 42회이다. 다음으로는 알운반을 목적으로 A000006 차량과 A000054 차량이 많이 방문했다. Sankey diagram의 결과로부터 A000007 차량과 A000006 차량, A000054 차량에 방역을 강화해야 한다는 것을 알 수 있다.

다음으로 Sankey diagram에 나타나지 않은 정보를 확인하기 위해 세 가지 방법론에 적용할 필요가 있다.

〈표 3〉 알고리즘 적용

	TF-IDF		방문 비율	
	차량번호	유사도	차량번호	유사도
1	A000054	0.25	A000054	1.00
2	A000007	0.17	A000007	0.44
3	A000006	0.15	A000006	0.23
4	A000049	0.51	A000049	0.14
5	A000029	0.43	A000069	0.08
6	A000069	0.02	A000029	0.07

〈표 3〉은 발생 농가를 TF-IDF 알고리즘과 방문 비율 알고리즘에 적용하여 유사도가 높은 차량을 뽑은 결과이다. 표의 왼쪽 부분이 TF-IDF 알고리즘 적용 결과이고, 오른쪽 부분이 방문 비율 알고리즘에 적용한 결과이다. Sankey diagram과 다른 점은 발생 농가와 A000054 차량의 유사도가 가장 높다는 점이다. A000054 차량의 방문 비율 알고리즘 결과를 보면 1의 값을 갖는다. 이 결과로부터 A000054 차량은 다른 시설을 방문하지 않는다는 것을 알 수 있다.

Sankey diagram을 통한 시각화의 결과와 같이 A000007 차량과 A000006 차량이 발생 농가와 관련성이 높았다. A000007 차량과 A000006 차량의 이동 경로를 기준으로 추가 연구가 필요하

다. 이를 위해 D2V 모델에 해당 차량의 동선을 input으로 넣어줬다.

<표 4> D2V 적용

	A000007		A000006	
	차량번호	유사도	차량번호	유사도
1	A000059	0.99	A000068	0.98
2	A000225	0.77	A000016	0.65
3	A000016	0.76	A000053	0.64
4	A000068	0.74	A000061	0.61
5	A000053	0.73		
6	A000061	0.70		

<표 4>는 D2V 모델을 이용하여 해당 차량의 경로와 유사한 경로를 가진 차량을 추출한 결과이다. 유사도가 0.6 이상인 차량을 추출하는 조건을 주었다. 표의 왼쪽 부분이 A000007 차량, 오른쪽 부분이 A000006 차량의 결과이다. A000068 차량, A000016 차량, A000053 차량, A000061 차량의 경우, A000006 차량과 A000007 차량의 경로와 모두 유사한 경로를 가지고 있다. <표 4>에 있는 차량은 발생 농가에 자주 방문하지 않거나 방문한 기록이 한 번도 없었을지라도 다른 시설에서 A000006 차량, A000007 차량과 만날 가능성이 높은 차량이다. 차량과 차량간 바이러스 전파의 가능성이 있으므로 D2V 모델에 의해 검출된 차량에 방역을 강화할 필요가 있다.

3.2 결과 해석

3.2.1 차량 방역 강화

Sankey diagram과 TF-IDF 알고리즘 및 방문 비율 알고리즘을 통해 발생 농가와 차량간 직접적인 연관성을 살펴볼 수 있었다. 발생 농가에 방문한 비율이 높을수록 차량이 HPAI 바이러스에 감염될 가능성이 높을 것이다. 해당 방법론을 통해 검출된 차량에 방역을 강화한다면

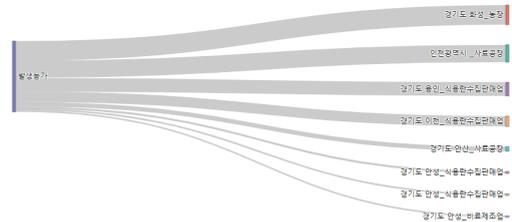
HPAI 바이러스 확산을 줄일 수 있을 것이다.

차량이 HPAI 바이러스에 감염된다면 차량과 차량간 전파도 고려해야 한다. 이를 위해 D2V 모델을 사용하여 차량간 이동 경로를 비교하였다. 발생 농가에 방문한 비율이 높은 차량의 이동 경로를 D2V 모델에 넣어줘 차량간 HPAI 바이러스 전파로 예상되는 차량을 검출할 수 있다. D2V 모델을 통해 검출된 차량에 방역을 강화한다면 HPAI 바이러스 확산을 줄일 수 있을 것이다.

비록 HPAI 바이러스에 감염된 농가에 가장 많이 방문한 차량이 HPAI 바이러스 유입의 가장 큰 원인이라는 가정이 추가된 결과이지만 향후 HPAI 바이러스가 검출됐을 때, 바이러스의 확산을 줄이기 위해 사용될 수 있을 것이다.

3.2.2 시설 방역 강화

Sankey diagram 을 이용하면 발생 농가와 연관성이 높은 시설 또한 알 수 있다. 바이러스와 연관성이 높은 시설의 방역을 강화한다면 HPAI 바이러스의 확산을 예방할 수 있을 것이다.



<그림 5> 시설별 빈도수 시각화

<그림 5>는 경기도 화성에 있는 박○준 농가에 방문한 차량이 이동한 시설과 박○준 농가의 관계를 보기 위해 전처리한 데이터를 Sankey diagram 을 이용해 시각화한 그림이다. 시설의 주소는 데이터의 보안상 생략하였다. 가독성을 위해 박○준 농가를 발생 농가로 정의하겠다. 발생 농가와 연관성이 가장 높은 시설은 경기도 화성에 있는 농장이다. 다음으로는 인천광역시

의 사료공장, 경기도 용인시와 이천시의 식용란 수집판매업이다. 해당 시설에 방역을 강화한다면 차량을 통해 시설에서 시설로의 HPAI 바이러스 전파를 줄일 수 있을 것이다.

3.3 추가 연구를 위한 제언

3.3.1 KAHIS 데이터의 한계

본 연구에 사용된 KAHIS 데이터는 보안 때문에 중간 경로가 생략된 정보만을 포함하였다. 중간 경로의 생략으로 인해 위험 차량이 아닐 것으로 예상되는 차량이 실제로 위험 차량이 아닌지 알 수 없다. 왜냐하면, 위험 지역은 HPAI에 감염된 농가 외에도 철새도래지 또는 철새도래지 인근 지역 등 많기 때문이다.

또한, KAHIS 데이터의 경우 GPS 단말기에 의해 차량의 출입정보를 수집, 관리하기 때문에 GPS가 꺼져있을 경우 데이터의 손실 문제가 발생한다.

3.3.2 연구 제언

더 정확한 연구를 위해 필요한 사항은 이렇다.

- 중간 경로가 포함된 KAHIS 데이터
- GPS 단말기 관리 또는 수정
- 지속적인 데이터 관리

축산차량은 축산차량등록제에 의해 차량의 출입정보를 수집, 관리한다. GPS 단말기가 이에 해당하는데 GPS 단말기를 사용했을 때의 단점은 단말기가 꺼진 상태로 이동하는 차량의 경우 해당 차량의 정보가 손실된다는 것이다. 이러한 단점을 보완하기 위해서는 GPS 단말기를 수시로 관리하거나 꺼지지 않는 GPS 단말기를 개발하고 적용하여 축산차량의 정보를 지속해서 수집하는 것이다. 이러한 방법을 통해 향후 더 많은 연구를 할 수 있을 것이다.

사 사

이 논문은 2018년도 정부(농림축산식품부)의 재원으로 농림식품기술기획평가원 지원을 받아 수행된 연구임.

참 고 문 헌

- [1] 농림축산검역본부 역학조사과. 『17/18 고병원성 조류인플루엔자 역학조사분석보고서』. 2018, 12, 김천: 농림축산검역본부 도서관.
- [2] 최대우, 주재운, 송유한 & 한예지, “CDR 자료를 이용한 고병원성 조류인플루엔자 분석”, 한국빅데이터학회지 제4권 제2호 pp. 13~22, 2019.

저 자 소 개



최 대 우(Dae-Woo Choi)

- 1986년: 서울대학교 계산통계학과 (학사)
- 1988년: 서울대학교 통계학과 (석사)
- 1994년: Rutgers University Statistics (박사)
- 1996년~현재: 한국외국어대학교 통계학과 교수
- 관심분야: 빅데이터 분석, 자연어 처리, 딥러닝 등



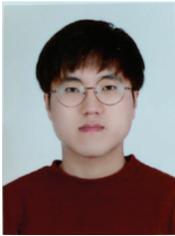
이 원 빈(Won-Been Lee)

- 2019년: 한국외국어대학교 물리학과 (학사)
- 2019년~현재: 한국외국어대학교 통계학과 (석사과정)
- 관심분야: 빅데이터 분석, seq2seq 자료 처리, 자연어 처리



송 유 한(Yu-Han Song)

- 2018년: 한국외국어대학교 수학과 (학사)
- 2018년~현재: 한국외국어대학교 통계학과 (석사과정)
- 관심분야: 머신러닝, 자연어 처리, 딥러닝 응용



강 태 훈(Tae-Hun Kang)

- 2019년: 한국외국어대학교 아랍어통번역학과 (학사)
- 2020년~현재: 한국외국어대학교 통계학과 (석사과정)
- 관심분야: 머신러닝, XAI, 딥러닝 응용



한 예 지(Ye-Ji Han)

- 2019년: 한양대학교 문화인류학과 (학사)
- 2019년~현재: 한국외국어대학교 통계학과 (석사과정)
- 관심분야: 빅데이터 분석, 자연어 처리, 텍스트 마이닝