

CDR 자료를 이용한 고병원성 조류인플루엔자 분석

The Analysis of HPAI Using CDR Data

최대우·주재윤[†]·송유한·한예지

한국외국어대학교 대학원 통계학과

요약

이 연구는 2018년도 정부(농림축산식품부)의 재원으로 농림식품기술기획평가원 지원을 받아 수행된 연구¹⁾이다. 고병원성 조류인플루엔자의 유입은 해외로부터 철새를 통해 유입되고 있으나 어떤 경로를 통해 감염 원인을 농장에 제공하는지 정확히 밝혀진 바 없다. 그리고 발생 농장으로부터 농장 간의 전이도 차량이 주 원인이라고 추정할 뿐, 전파 주 원인이 정확히 밝혀진 것은 아니다.

본 논문 연구에서는 KT가 제공하는 CDR(Call Detailed Record) 데이터를 기반으로, 발생후보지로 추정되는 철새도래지를 방문한 사람이 어떤 경로를 거쳐 감염 농장으로 유입이 되는지 해당 시기의 인구 흐름을 보려고 한다.

■ 중심어 : CDR(Call Detailed Record), HPAI(고병원성 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. The inflow of highly pathogenic avian influenza is coming through migratory birds from abroad, but it is not known exactly what pathways provide the farm with the cause of the infection. And the transition between farms from the generated farms only assumes that the vehicle is the main cause, and the main cause of the spread is not exactly known.

Based on the call detailed records (CDR) data provided by KT, the study aims to see how people visiting migratory bird-watching sites, presumed to be the site of the outbreak, will flow through infected farms.

■ Keyword : CDR(Call Detailed Record), HPAI(Highly Pathogenic Avian Influenza)

I. 서론

1.1 연구의 목적

고병원성 조류인플루엔자의 유입은 해외로부터 철새를 통해 유입되고 있으나 어떤 경로를 통해 감염 원인을 농장에 제공하는지 정확히 밝혀진 바 없다. 그리고 발생 농장으로부터 농장 간의 전이도 차량이 주원인이라고 추정할 뿐, 전파 주원인이 정확히 밝혀진 것은 아니다.

본 논문 연구에서는 KT가 제공하는 CDR (Call Detailed Record) 데이터를 기반으로, 발생후보지로 추정되는 철새도래지를 선별하여 방문한 사람들의 경로를 확인한다. 경로들을 확인했을 때 많이 나온 구간들이 해당 시기에 인구의 흐름을 반영한 것으로 판단할 수 있다. 해당 인구의 흐름을 통해 바이러스의 전파가 어느 방향으로 이루어졌는지 확인하고자 한다.

II. 본론

2.1 CDR 자료에 대해

2.1.1 CDR 자료 설명

고위험성 조류인플루엔자(HPAI) 발생 농장의 발병 요인을 찾는 것을 목표로 발생의 원인으로 추정되는 야생 조류에 의해 오염된 지역에서 어떠한 경로를 통해 전파가 되었는지 확인하기 위해 사람 이동 자료를 이용하였다. 사람 이동 자료는 현 과제의 협력기관인 KT에서 제공하는 CDR 자료이다.

CDR 자료는 KT에서 운영하는 기지국에 반경 내 통신사 가입자가 들어왔을 때 기지국에서 시그널을 받아 해당 사람이 머무른 시간과 그 다음 기지국으로 이동했을 때 다음 기지국의 정보를 전달해주는 sequence to sequence 자료이다.

KT에서 발생 농장별로 CDR 자료를 구성하는 것은 고위험성 조류인플루엔자 발생 농장의 신고 일자에서 과거 7일간 농장을 방문한 사람들의 과거 경로들을 추출하였다. 제공받은 자료는 기지국으로 들어온 시간, 기지국에서 나간 시간, 비식별화된 사람 정보, 비식별화된 기지국 정보, 읍·면·동 단위까지의 기지국 주소 정보를 이용하였다.

KT가 제공해주는 CDR 자료의 컬럼들의 정보는 다음과 같다.

- 기지국 비식별정보
- 인구 비식별 일련번호
- 인구의 기지국 위치 시간대
- 기지국 주소 정보(시도, 시군구, 읍면동)

2.1.2 CDR 자료 전처리

위 자료를 분석하고 지도상에 표시하기 위해서는 기지국 마다 위치 정보가 필요로 하나, 기지국 정보는 공개하기 어려운 자료이기에 정확한 위치는 알 수 없었다. 기지국의 읍·면·동 단위까지의 주소를 이용하여, 읍·면·동 단위의 중심 좌표를 기지국의 좌표로 연구를 위해 활용을 했다. 기지국 주소의 중심 좌표 정보를 이용하여 기지국이 철새도래지에 속하는지를 확인하고 정리를 하였다.

위에서 이야기한 내용들에 대해서 다음과 같이 정리하였으며, 변수들의 이름과 설명은 다음과 같다.

- address : 기지국 주소 정보
- lat : address 변수의 주소의 중심 위도
- long : address 변수의 주소의 중심 경도
- new_bts_id : address 변수의 주소에 따른 새로운 기지국 라벨
- manage_no : address 변수의 주소가 어떤 철새도래지에 속하는지에 대한 값

2.2 발생후보지 선별 및 CDR 자료 활용 방법

CDR 자료를 활용하여 발생후보지로부터 발생 농장 또는 기지국으로 온 경우는 철새도래지에서 어떻게 왔는지, 철새도래지는 어떠한 기준으로 선별을 했는지 설명을 하고자 한다.

2.2.1 발생 후보지 선별 과정

발생후보지 선별은 철새도래지와 야생조류에서의 HPAI 바이러스 항원 검출 자료를 바탕으로 구성을 하였다.

- 1) 발생 농장에서의 항원의 바이러스 타입을 확인을 진행한다.
- 2) 발생 농장에서의 바이러스 타입과 맞는 야생조류 HPAI 바이러스 항원 검출지점을 확인한다. 1번과 2번 과정을 거치는 이유는 바이러스 타입이 같아야 발생 원인의 근거로써 활용을 할 수 있기 때문이다.
- 3) 항원 검출 지점은 발생 기지국에 속한 농장의 발생 일자를 고려해서 선별을 하였다. 3번 과정은 다음과 같은 과정을 통해서 진행을 하였다.

3.1) 발생 농장의 발생 일자가 해당 월의 15일 기준으로 초반 및 후반으로 구분을 진행한다.

3.2.1) 해당 월의 15일 기준으로 초반일 경우에는 전월의 15일부터 해당 월의 30일까지의 항원 검출 지점을 선별한다. 예를 들면, 12월 10일 발생은 11월 15일부터 12월 30일까지의 항원 검출 지점들을 고려한다.

3.2.2) 해당 월의 15일 기준으로 후반일 경우에는 해당 월의 1일부터 다음 월의 15일까지의 항원 검출 지

점을 선별한다. 예를 들면, 12월 17일 발생은 12월 1일부터 1월 15일까지의 항원 검출 지점을 선별한다. 위와 같은 기준으로 선정한 이유는 항원 검출 시기가 발생 농장의 발생 시기보다 늦은 경우와 HPAI 바이러스 생존 시기를 크게 잡고 고려를 한 것이다.

- 4) 3번 이후 항원 검출 지점에서 항원을 배출한 철새 종을 확인하여, 해당 종이 서식하고 있는 철새도래지들을 발생후보지로 선별을 하였다.

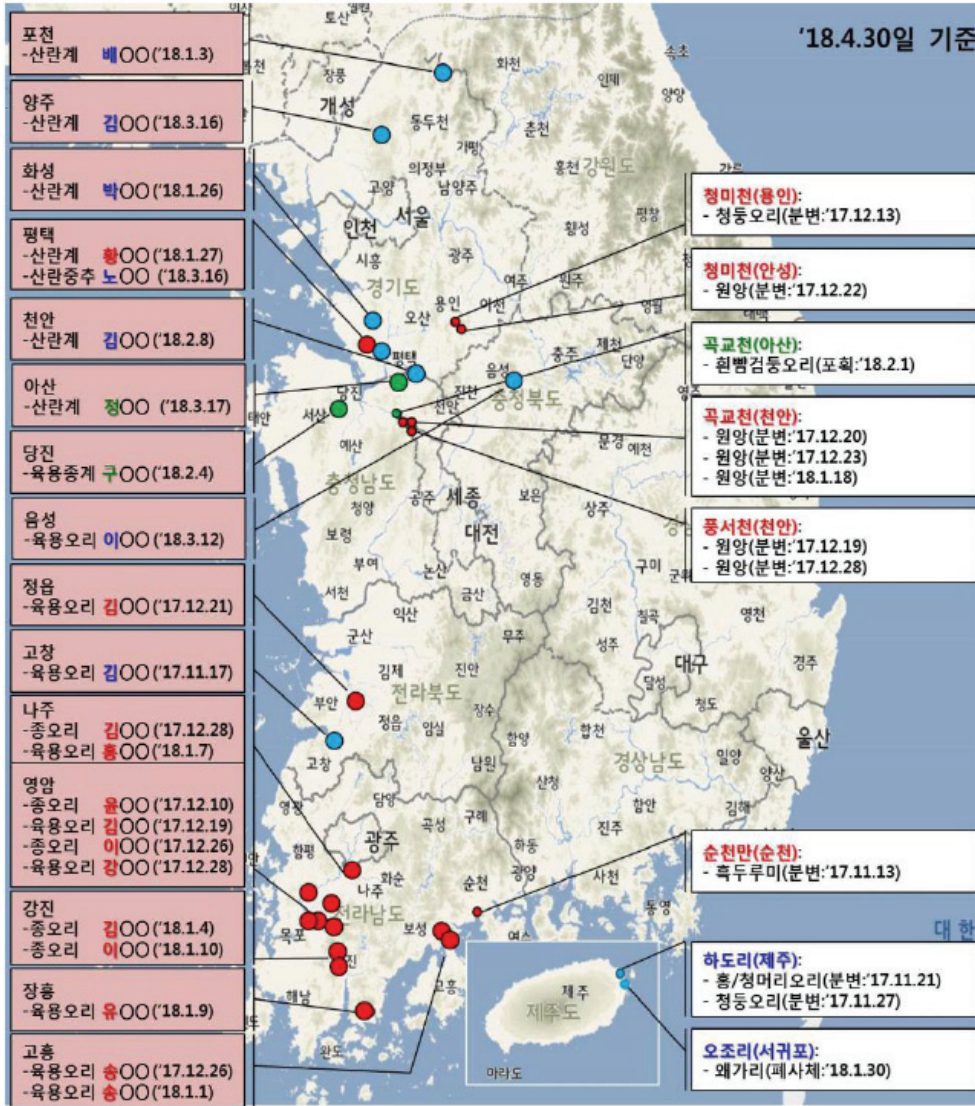
2.2.2 CDR 자료 활용 방법

발생후보지 선별이 진행된 다음 발생 기지국 또는 농장까지의 경로들은 다음과 같은 정제 과정을 거쳤다.

- 1) 발생 후보지로 선별한 철새도래지에 10분 이상 머물고 발생 기지국에 10분 이상 머물 경우에 대한 경로들을 추출을 하였다.
- 2) 철새도래지를 방문한 이후 다른 철새도래지를 지나가는 경우가 있어도 발생 기지국으로 갈 때까지의 경로를 찾았다. 추가 설명을 하면 기준을 잡은 철새도래지에서 다른 철새도래지로 가게 되는 경우를 무시하는 이유는 다른 철새도래지도 경로를 찾는 알고리즘에서 이후에 기준이 되기에 기준이 잡은 철새도래지에서 발생 기지국으로 간 경우만 확인을 하였다.
- 3) CDR 자료에서 사람들의 이동 경로는 발생 농장의 발생 일자의 15일 전부터 기지국까지의 온 경로들을 확인을 하였다.

<그림 1>은 농림축산검역본부에서 작성한 역학조사보고서에 추출한 그림²⁾이며, 위 그림을 통

2) 농림축산검역본부(2018.12), '17/18 고병원성 조류인플루엔자 역학조사분석보고서(19쪽)



〈그림 1〉 - 발생농장 및 야생조류 바이러스 타입 별 HPAI 검출 분포

해서 발생 농가의 바이러스 타입과 야생 조류 분변에서 검출된 바이러스 타입과의 비교를 진행을 하였다.

3.1 바이러스 별 최초 발생 농장에 대한 분석

3.1.1 고창형 바이러스- 2017년 11월 17일 발생

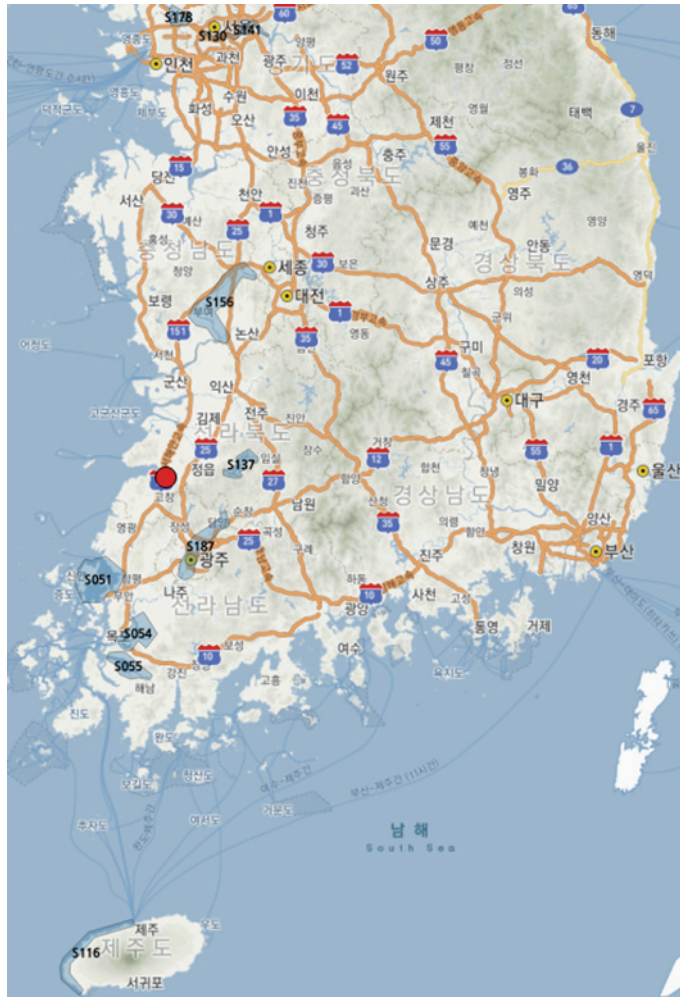
2017년 11월 17일 처음으로 고창 지역의 오리 농장에서 고창형 바이러스가 검출이 되었다.

17/18년도 최초로 농장에서 고병원성 조류인플루엔자가 검출되었으며, 해당 바이러스는 야생 조류 분변에서의 바이러스가 이동 매개체를 통해서 전파가 되었을 가능성이 크다. 발생 후보지로 추정되는 철새도래지로부터 사람들의 이동이 발생 농장에 대한 경로를 확인을 진행하였다.

발생 후보지 선별을 위해 우선 발생 일자와 바이러스 타입과 동일한 야생 조류 항원 검출 일자를 확인을 하였다. 흑두루미, 청둥오리가 11월,

[표 1] - 발생 농장까지 경로 중 많은 경로를 배출한 철새도래지 10개

Manage_no	Count	Migratory_nm
S187	849	영산강 상류(승촌보 상류-담양호 하류)
S130	148	성산~성수대교
S141	120	성수대교~팔당댐
S054	115	영산호
S156	111	금강중류(황산대교~청벽대교)
S055	108	영암호
S178	106	한강하류(일산대교~행주대교)
S051	104	무안군 현경면,운남면
S116	101	용담~대정 해안
S137	71	옥정호



<그림 2> - 상위 10개 철새도래지 위치 및 11월 17일 발생 농장 위치

12월 서식하고 있었던 철새도래지 83개를 동시 센서스 자료를 통해 발생후보지로 선별을 하였다. 11월 2일부터 11월 17일까지 15일간 발생 후보지로부터 발생 농장까지의 경로를 찾았으며, 결과는 다음과 같다.

- 1) 발생후보지로 선정된 철새도래지 중 가장 많은 경로를 보인 철새도래지 확인

CDR 자료를 통해 발생 후보지로 선정된 83개의 철새도래지에서 발생 농장까지의 경로를 확인

하였을 때, 77개의 철새도래지에서만 발생 농장까지의 경로를 확인할 수 있었다. 위 표와 그림은 철새도래지 중에서도 11월 17일 발생농장으로 가장 많이 가는 것으로 확인되어진 철새도래지의 상위 10개를 표시한 것이다. 많이 확인된 철새도래지는 영산강 상류이며, 발생 농장을 중심으로 남쪽에 위치하여 해당 시기에 철새도래지를 지난 이후 인구 흐름이 남에서 북으로 될 것이라는 추측을 할 수 있다. <그림 2>에서 빨간 동그라미가 발생농장 위치, polygon 형태로 표시가 되어 있는 것이 철새도래지이다.

<표 2> - 발생 농장을 중심으로 한 유입 구간 빈도수 상위 25개

route	count	from_address	to_address
NB14682 → NB14670	1069	전라북도 고창군 부안면	전라북도 고창군 신림면
NB14684 → NB14670	977	전라북도 고창군 성내면	전라북도 고창군 고창읍
NB14666 → NB14670	845	전라북도 고창군 신림면	전라북도 고창군 성내면
NB14705 → NB14670	720	전라북도 부안군 줄포면	전라북도 김제시 금구면
NB14670 → NB14682	645	전라북도 고창군 흥덕면	전라북도 정읍시 소성면
NB14670 → NB14684	500	전라북도 고창군 흥덕면	전라북도 부안군 줄포면
NB14670 → NB14705	354	전라북도 고창군 흥덕면	전라북도 고창군 부안면
NB14670 → NB14666	342	전라북도 고창군 흥덕면	전라북도 정읍시 감곡면
NB14664 → NB14670	179	전라북도 고창군 고창읍	전라북도 부안군 부안읍
NB14693 → NB14670	166	전라북도 정읍시 소성면	전라북도 고창군 고창읍
NB14702 → NB14670	86	전라북도 부안군 보안면	전라북도 정읍시 용계동
NB14670 → NB14693	66	전라북도 고창군 흥덕면	전라북도 부안군 주산면
NB14670 → NB14664	55	전라북도 고창군 흥덕면	전라북도 정읍시 농소동
NB14689 → NB14670	51	전라북도 고창군 고수면	전라북도 고창군 부안면
NB14675 → NB14670	44	전라북도 고창군 아산면	전라북도 정읍시 수성동
NB14670 → NB14702	40	전라북도 고창군 흥덕면	전라북도 고창군 고수면
NB14708 → NB14670	32	전라북도 부안군 주산면	전라북도 정읍시 공평동
NB14691 → NB14670	25	전라북도 고창군 대산면	전라북도 부안군 보안면
NB6586 → NB14670	24	전라북도 정읍시 용계동	전라북도 정읍시 농소동
NB875 → NB14670	20	전라북도 정읍시 수성동	전라북도 완주군 이서면
NB14678 → NB14670	18	전라북도 부안군 부안읍	전라북도 정읍시 농소동
NB14670 → NB14675	17	전라북도 고창군 흥덕면	전라북도 정읍시 연지동
NB14662 → NB14670	14	전라북도 정읍시 입암면	전라북도 정읍시 연지동
NB1677 → NB14670	14	전라북도 정읍시 농소동	전라북도 고창군 고수면
NB14670 → NB14689	12	전라북도 고창군 흥덕면	전라북도 정읍시 수성동

2) 발생 농장을 중심으로 한 유입 구간 확인

<그림 3>을 통해 발생 후보지에서 발생 농장까지의 총 2749개의 경로 중에서 발생 농장의 유입 방향을 확인하였다. <표 2> 결과와 같이 확인하였을 때, NB14682 → NB14670, NB14684 → NB14670, NB14666 → NB14670, NB14705 → NB14670 구간에서 유입이 많이 보였다. 즉, 서쪽, 동쪽, 남동쪽, 북서쪽 방향 순으로 많이 유입이 된 것을 확인할 수 있다. 다음 순위는 위 결과의 역방향들이었으며 발생 농장을 중심으로 기지국간에 양방향 순위를 확인을 했을 때 NB14682

- NB14670, NB14684 - NB14670, NB14666 - NB14670, NB14705 - NB14670 순으로 인구 유동이 많았던 것을 확인할 수 있다.

3) 발생 농장 이외의 인구 흐름에 대한 구간 확인

<표 3>은 2749개의 철새도래지에서 발생 농장까지의 경로 내에서 가장 많이 나온 구간과 그 구간을 차단했을 때 2749개의 경로 중 몇 개의 경로를 차단할 수 있는지 확인한 표이다. 발생 농장으로 들어가는 구간들을 제외하고 결과를 추출하였다. <표 3>의 변수들의 설명은 다음과 같다.



<그림 3> - <표 2> route 변수들을 지도상 표현

〈표 3〉 - 구간 중 발생 농장 제외한 구간 차단 시 차단율 상위 25개

block_route	block count	block rate	count	from_address	to_address
NB14664 → NB14666	883	32.121	1149	전라북도 고창군 고창읍	전라북도 고창군 신탐면
NB14689 → NB14664	734	26.701	1013	전라북도 고창군 고수면	전라북도 고창군 고창읍
NB14693 → NB14684	630	22.917	778	전라북도 정읍시 소성면	전라북도 고창군 성내면
NB14672 → NB14688	533	19.389	731	전라북도 완주군 이서면	전라북도 김제시 금구면
NB6586 → NB14693	527	19.171	623	전라북도 정읍시 용계동	전라북도 정읍시 소성면
NB14702 → NB14705	499	18.152	590	전라북도 부안군 보안면	전라북도 부안군 줄포면
NB14666 → NB14682	495	18.007	564	전라북도 고창군 신탐면	전라북도 고창군 부안면
NB14688 → NB14734	429	15.606	496	전라북도 김제시 금구면	전라북도 정읍시 감곡면
NB14772 → NB14678	395	14.369	479	전라북도 부안군 동진면	전라북도 부안군 부안읍
NB14666 → NB14664	374	13.605	545	전라북도 고창군 신탐면	전라북도 고창군 고창읍
NB7816 → NB6586	363	13.205	432	전라북도 정읍시 공평동	전라북도 정읍시 용계동
NB14678 → NB14708	357	12.987	409	전라북도 부안군 부안읍	전라북도 부안군 주산면
NB876 → NB1677	346	12.587	411	전라북도 정읍시 하북동	전라북도 정읍시 농소동
NB14675 → NB14682	338	12.295	367	전라북도 고창군 아산면	전라북도 고창군 부안면
NB1677 → NB875	334	12.150	468	전라북도 정읍시 농소동	전라북도 정읍시 수성동
NB14692 → NB14689	319	11.604	415	전라북도 고창군 성송면	전라북도 고창군 고수면
NB1677 → NB7816	310	11.277	350	전라북도 정읍시 농소동	전라북도 정읍시 공평동
NB14705 → NB14702	298	10.840	394	전라북도 부안군 줄포면	전라북도 부안군 보안면
NB875 → NB1677	293	10.658	443	전라북도 정읍시 수성동	전라북도 정읍시 농소동
NB14688 → NB14672	288	10.477	464	전라북도 김제시 금구면	전라북도 완주군 이서면
NB123 → NB1677	266	9.677	344	전라북도 정읍시 연지동	전라북도 정읍시 농소동
NB1677 → NB123	253	9.203	354	전라북도 정읍시 농소동	전라북도 정읍시 연지동
NB875 → NB123	237	8.621	389	전라북도 정읍시 수성동	전라북도 정읍시 연지동
NB14664 → NB14689	223	8.112	403	전라북도 고창군 고창읍	전라북도 고창군 고수면
NB123 → NB875	214	7.784	381	전라북도 정읍시 연지동	전라북도 정읍시 수성동

- block_route : 차단 구간
- block_count : 발생 후보지에서 발생 농장까지의 총 경로 개수 중 차단 개수
- block_rate : 발생 후보지에서 발생 농장까지의 총 경로 개수 중 차단 개수의 비율
- count : 발생 후보지에서 발생 농장까지의 총 경로 내에서 해당 구간이 나온 빈도수
- from_address : block_route 값에서 ‘→’ 기준으로 왼쪽 기지국의 읍, 면, 동 주소
- to_address : block_route 값에서 ‘→’ 기준으

로 오른쪽 기지국의 읍, 면, 동 주소

<그림 4>는 <표 3>에 나온 block_route의 구간들을 QGIS 프로그램을 이용하여 지도상에 기지국 간에 최단 거리를 매핑을 했다. 해당 시기에 인구의 흐름이 NB14689 → NB14664 → NB14666 기지국 순으로 즉, 남쪽에서 북쪽으로 이동을 하는 것을 알 수 있다. 남쪽에서는 경로 시작 지점 중 가장 많이 확인된 S187(영산강 상류) 철새도래지가 있으며 인구 유동이 밑에서 위

로 가지 않을까 앞에서 추측을 하였다. 그 다음의 경로들을 확인을 하면 발생 농장 북서쪽 방향인 NB14693 → NB14684, NB14672 → NB14688에서 인구의 흐름이 있었다는 것을 알 수 있다.

최초 발생 농장이며, 처음 고창형 바이러스가 발견된 농장에 대해 CDR 자료로 인구의 흐름을 확인했을 때, 발생 농장 중심으로 남쪽 방향과 북서쪽 방향에서 많이 인구의 흐름이 있어서 그러한 인구 흐름들 사이에서 농장들이 많이 위치를 하는 곳이 있다면 그 이전에 방역 초소 등을 세워야 하지 않을까 하는 추가적인 연구가 필요로 할 것 같다.

III. 결론

3.1 CDR 자료 활용에 대한 결론

현재 대부분의 역학 분석은 차량에 의해서 이뤄졌으며, 인구 흐름을 활용한 분석은 많이 이뤄지지 않았다. 이번 분석을 통해 고병원성 조류인플루엔자를 분석을 인구 흐름을 통해 발생 이전의 인구 흐름들을 알 수 있었다. 그러한 흐름들을 통해 어느 구간에서 차단을 했을 때 효과가 있을지 생각할 수 있게 해주는 부분들이 있다. 이번에는 17/18년도 바이러스 별 초기 발생을 했을 당시 인구 흐름들을 파악을 해보았다. 이 논문에 목적은 CDR 자료를 통해 고병원성 조류인플루엔자가 발생했을 당시의 사람들의 이동 경로 및 전체적인 인구의 흐름들을 파악을 하는 것이었으며, 그러한 것들이 가능하다는 것을 확인할 수 있었다. 현 방법 이외에도 다양한 방법을 적용할 수

있다는 것을 알 수 있었다.

3.2 최초 발생농장에 대한 결과

17/18년에 발생한 최초 바이러스인 고창형 바이러스에 대한 CDR 자료를 통해 인구 흐름을 확인을 하였다. 인구 흐름들을 확인했을 때에는 주변 철새 도래지 근처에서 많았다는 것을 알 수 있다.

고창형 바이러스는 발생 농장 중심으로 남쪽 방향과 북서쪽 방향에서 많이 인구의 흐름이 있었으며, 이를 차량의 이동과 접촉해서 보게 된다면 더 유의미한 결과를 볼 수 있을 것 같으며, 또는 비발생에 관련된 자료와 비교를 통해서 유의미한 결과를 확인할 수 있을 것이라 보인다. 이번 논문에서는 역학분석에서 CDR 자료를 접목시키는 것이 가능한지를 확인하는 연구하는 것을 목표로 진행했으며 가능함을 확인하였다.

참 고 문 헌

- [1] 농림축산검역본부 역학조사과. 17/18 고병원성 조류인플루엔자 역학조사분석보고서. 2018, 12, 김천: 농림축산검역본부 도서관.

사 사

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

저자 소개



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



주 재 윤(Jae-Yun Joo)

- 2018년 : 한국외국어대학교
통계학과 (학사)
- 2019년~현재 : 한국외국어
대학교 통계학과 (석사과정)
- 관심분야 : 빅데이터 분석,
seq2seq 자료 처리, XAI



송 유 한(Yu-Han Song)

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



한 예 지(Ye-Ji Han)

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