한국빅데이터학회지
제5권 제2호, 2020, pp. 69-76 https://doi.org/10.36498/kbigdt.2020.5.2.69

고위험성 조류인플루엔자(HPAI) 확산 방지를 위한 GAN 기반 가상 데이터 생성*

Generating GAN-based Virtual data to Prevent the Spread of Highly Pathogenic Avian Influenza(HPAI)

최대우[1] · 한예지[2†] · 송유한[2] · 강태훈[2] · 이원빈[2]

한국외국어대학교 자연과학대학 통계학과 교수[1], 한국외국어대학교 대학원 통계학과[2]

요 약

이 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구이다.[1)] 고병원성조류인플루엔자(Highly Pathogenic Avian Influenza, HPAI)는 병원성이 높은 조류인플루엔자 바이러스 감염에 의하여 발생하는 조류의 급성 전염병으로 닭, 오리 등 가금류에서 피해가 심각하게 나타난다. 고병원성 조류인플루엔자(HPAI)는 연중으로 발생하기보다는 겨울철에 집중하여 발생되는 양상을 보이며, 특정 기간에는 아예 발생하지 않는 경우가 있다. 이와 같은 HPAI의 특성으로 인해 충분한 양의 실제 데이터가 축적되지 못하는 문제점이 있다. 본 논문 연구에서는 GAN 네트워크를 활용하여 결측치를 포함하고 있는 실제와 유사한 데이터를 생성하였으며 해당 과정을 소개한다. 본 연구 결과는 HPAI가 발생하지 않은 특정 시기에 대하여 실제와 유사한 시뮬레이션 데이터를 생성하여 위험도를 측정하는데 이용될 수 있다.

■ 중심어 : GAN(Generative Adversarial Network), HPAI(고위험성 조류인플루엔자), 가상 데이터 생성

Abstract

This study was conducted with the support of the Information and Communication Technology Promotion Center, funded by the government (Ministry of Science and ICT) in 2019. Highly pathogenic avian influenza (HPAI) is an acute infectious disease of birds caused by highly pathogenic avian influenza virus infection, causing serious damage to poultry such as chickens and ducks. High pathogenic avian influenza (HPAI) is caused by focusing on winter rather than year-round, and sometimes does not occur at all during a certain period of time. Due to these characteristics of HPAI, there is a problem that does not accumulate enough actual data. In this paper study, GAN network was utilized to generate actual similar data containing missing values and the process is introduced. The results of this study can be used to measure risk by generating realistic simulation data for certain times when HPAI did not occur.

■ Keyword : GAN(Generative Adversarial Network), HPAI(Highly Pathogenic Avian Influenza), Simulation Data Generation

2020년 11월 16일 접수; 2020년 12월 04일 수정본 접수; 2020년 12월 16일 게재 확정.
* 이 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임.
† 교신저자 (dowlhan95@gmail.com)
1) 2018-0-00430, 인공지능기술을 활용한 고병원성 조류인플루엔자 국내 유입, 발생 조기 감지 및 확산 대응시스템 개발

Ⅰ. 서 론

고병원성 조류인플루엔자가 최초로 농가에 발생한 후에 다른 농장으로 확산하는 속도는 빠르며 높은 폐사율과 산란율 저하 등으로 인한 사회적, 경제적 손실이 막대하다. 그렇기에 HPAI 발생 시 주변 지역으로의 전염을 신속하게 차단해야 한다.

최근에는 AI(Artificial Intelligence)를 이용하여 HPAI 발생 시 주변 지역의 위험도를 예측하고, 위험도에 따른 방역 조치가 신속하게 이루어질 수 있도록 하는 다양한 예측 모델들이 개발되고 있다. 이러한 예측 모델들은 주로 머신러닝을 기반으로 하기에 예측 모델들의 정확도 및 유효성은 훈련 데이터에 의존하게 된다.

HPAI가 특정 농장에서 발생하면 HPAI의 감염 확산을 방지하기 위해 지정 범위 내 모든 가금류를 폐기하게 시킨다. 폐기된 사항은 HPAI 실제 데이터에 결측 데이터로 표시된다. 또한 HPAI는 연중으로 발생 되기보다는 겨울철에 집중하여 발생 되는 양상을 보이며, 특정 기간에는 발생하지 않는 때도 있다. 그렇기에 예측 모델의 학습을 위한 충분한 양의 실제 데이터가 축적되지 못한다.

위와 같은 HPAI 특성을 고려하였을 때 실제 데이터와 유사한 가상 데이터를 생성하여 예측 모델의 학습에 필요한 충분한 양의 데이터 확보가 필요하다. 충분한 데이터 확보를 위해 GAN을 통해 가상 데이터를 생성하였다. 더욱 정확한 시뮬레이션을 위해서는 실제 데이터에 결측 데이터가 반영된 것과 같이 가상 데이터에도 결측 데이터가 반영되어야 한다. 이는 GAN을 통해서 HPAI 발생 시기와 위치를 반영하여 결측치를 포함하는 가상 데이터를 생성할 수 있다.

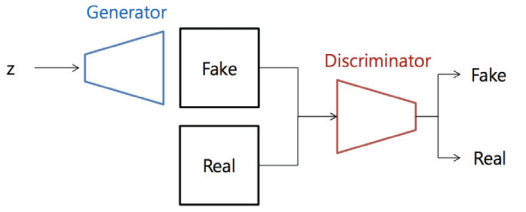
본 연구에서는 HPAI의 특성으로 인해 야기되는 충분하지 않은 데이터의 양과 데이터 내의 결측치 문제를 해결하기 위해 실제와 유사한 가상 데이터를 GAN(Generative Adversarial Network)을 통해 생성하였다. 해당 연구를 통해 실제 데이터와 유사하게 특정 조건을 부여하여 결측치까지 포함한 가상 데이터를 생성하는 과정을 확인하고 이를 통해 충분한 양의 데이터를 확보한 뒤 예측 모델의 정확도를 높이는데 기여할 수 있는 방안을 제시하고자 한다.

Ⅱ. 연구 방법

2.1 GAN

2.1.1 GAN 소개

GAN(Generative Adversarial Network)은 이미지 생성이나 영상 생성 등 데이터 생성 분야에서 활발하게 연구되고 있는 알고리즘이다. 이는 생성적 적대 신경망 알고리즘으로 Generator와 Discriminator 2개의 네트워크로 이루어져 있고, 두 네트워크가 상반되는 목적으로 경쟁을 하면서 학습을 하여 합리적인 가상 데이터를 생성할 수 있다.



〈그림 1〉 GAN(Generative Adversarial Network)

<그림 1>을 통해 보았을 때, Generator 네트워크는 생성된 z를 받아서 실제 데이터와 비슷한 데이터를 만들어내도록 학습한다. 그리고 Discriminator 네트워크는 실제 데이터와 Generator가 생성한 가상 데이터를 구별하도록 학습한다.

GAN 알고리즘의 특징을 이용하여 임의의 분포를 가정한 모형에서 학습된 잠재 변수(latent vector)에 날짜 및 공간 조건을 반영하여 가상 데

이터를 생성하였다. 추가적으로 특정 농장에서 HPAI가 발생하게 되면, 다른 농장으로의 감염 확산을 방지하기 위해 검역 본부가 지정한 범위 내에 있는 모든 가금류를 폐기시킨다. 그렇기에 HPAI 관련 실제 데이터에는 결측 데이터가 다수 포함이 되어 있다. 즉, 더욱 정확한 시뮬레이션 데이터 생성을 위해서는 실제 데이터에 결측 데이터가 반영되어있는 것과 같게 가상 데이터 생성 시에도 결측 데이터가 포함되어야 한다.

이와 같은 HPAI의 조건적인 특징을 반영하여 본 연구에서는 실제와 유사한 형태의 가상 데이터를 GAN 시뮬레이션을 통하여 생성하였다.

2.1.2 GAN 모델 구조

GAN 은 Generator 와 Discriminator 2 개의 네트워크 구조로 이루어져 있으며 두 네트워크가 상반되는 목적으로 경쟁을 하면서 학습을 하는 구조이다. GAN 을 고안한 Ian Goodfellow 는 이러한 구조를 minimax two-player game 과 같다고 설명한다.

$$\min_G \max_D V(D,G) = E_{x \sim p_{data}(x)}(\log D(x)) + E_{z \sim p_z(z)}(\log D(G(z)))$$

<식 1> The structure of GAN

식(1)에서 minimax란 추정되는 최대의 손실을 최소화하는 기법이다. 이는 최악의 경우를 발생하게 하는 손실을 최소화한다는 규칙을 의미한다.

Discriminator는 Generator의 생성물과 실제 데이터를 잘 구별하는 것을 목적으로 하는 binary cross-entropy를 손실 함수로 하여 분류를 최대화하는 것을 목적으로 학습한다.

Generator 손실 함수의 경우 생성물을 Discriminator Network에 보낼 때 실제 데이터로 판단하게 하여 손실을 최소화하는 목적으로 학습한다. 다음은 Discriminator 및 Generator Network의 손실 함수를 수식화한 것이다.

$$\max_D V(D) = E_{x \sim p_{data}(x)}(\log D(x)) + E_{z \sim p_z(z)}(\log 1 - D(G(z)))$$

<식 2> Discriminator 손실 함수

$$\min_G V(G) = E_{z \sim p_z(z)} \log(1 - D(G(z))) - \frac{1}{2} E_{z \sim p_z(z)} \log(D(G(z)))$$

<식 3> Generator 손실 함수

2.2 GAN 학습 데이터

2.2.1 GAN 학습 사용 데이터

GAN 학습에 사용되는 데이터는 기존 예측 모형에 사용된 실제 데이터를 활용하여 이에 특정 조건을 주어 만들어진 데이터이다. 그렇기에 기존 예측 모형에 사용된 실제 데이터에 대한 이해가 필요하다.

실제 데이터는 총 612,906개의 행으로 이루어진 다이나믹 데이터 마트이다. 다이나믹 데이터 마트는 분석 단위인 농장에 따라 시간 및 공간 등 다양한 정보를 이용한 데이터 마트이다. 해당 데이터는 HPAI 확산 예측 모델 학습에 사용되는 데이터로 농장, 차량, 날씨, 공간, 철새, 방역 총 6가지 카테고리에 대한 변수들로 구성되어 있다. 데이터 마트의 각 변수들은 국가가축방역시스템인 KAHIS 자료 및 다른 협력 기관들이 제공한 자료를 활용하여 구성되어 있다. 해당 데이터는 HPAI가 발생하였을 때 발생 일자 이후에 발생 농장에 대한 정보들은 사라지기에 특정 발생 기간 이후에 값들은 결측치로 처리되어 있다는 특징을 갖고 있다. 그렇기에 이후 GAN으로 가상 데이터를 생성할 시에 결측치에 대한 데이터도 함께 생성해줘야 한다.

2.3 GAN 학습 데이터 전처리

2.3.1 학습 데이터를 위한 전처리

GAN 학습을 위해 사용된 데이터는 기존 예측 모형에 사용된 실제 데이터 중 농장, 차량, 날씨, 방역 변수의 일부를 활용하며 해당 변수들은 numerical 한 변수와 categorical 한 변수들로 이루어져 있다. 각 변수는 척도가 다르기에 이들을 모델 학습 데이터로 사용하기 위해서는 표준화 과정을 거쳐야 한다. GAN 학습에 사용된 데이터는 변수별 값들을 0과 1 사이 범위의 값들로 먼저 scaling 해주는 min-max scaling을 활용하였고 여기서는 최대, 최솟값을 -1과 1 사이로 범위를 넓혀 전처리를 진행하였다.

2.3.2 Numerical & Categorical 변수 scaling

학습에 사용될 변수만 선택하여 그 중 numerical 변수와 categorical 변수로 분리하여 지정해준다. 이후 numerical 변수들에 대해서는 -1과 1 사이의 값으로 min-max scaling을 해준다. Categorical 변수들에 대해서는 one-hot encoding을 한 후 dummy variable 형태로 변경하여 값을 넣어준다. 이처럼 numerical 및 categorical 변수에 대하여 min-max scaling을 각각 진행한 후 numerical 변수와 categorical 변수를 하나의 데이터로 만들어 준다.

2.3.3 날짜 및 위치 조건 데이터 생성

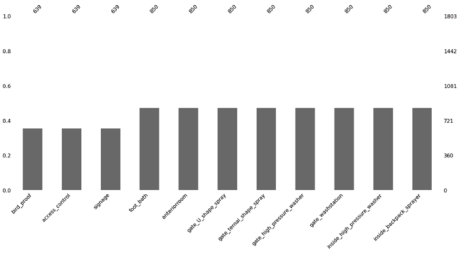
파이썬에서 시간을 처리하는 패키지인 datetime을 활용하여 이후 GAN 모델 학습 시 날짜 조건을 부여하기 위해 날짜 데이터를 생성한다. 날짜 데이터는 총 1년 치의 데이터로 1월 1일부터 12월 31일까지의 일자 총 1년에 해당하는 일자별로 min-max scaling을 하여 -1과 1 사이의 값을 붙여서 데이터를 생성한다.

다음으로는 GAN 모델 학습 시 위치 조건을 부여하기 위해 위치 데이터를 생성한다. 우리나라 위, 경도에 해당하는 범위를 지정하여 matrix를 만들어 각 위, 경도에 대하여 날짜와 동일하게 min max scaling을 하여 -1과 1 사이의 값을 붙인 후 데이터를 생성한다.

2.3.4 결측치 패턴 확인 및 label encoding

다음으로는 실제 데이터를 활용하여 날짜 및 위치 조건을 주어 전처리한 GAN 학습 데이터 내의 결측치 패턴을 확인한다. 해당 과정을 통해 실제 데이터와 유사한 형태로 결측치를 포함하여 가상 데이터를 생성할 수 있다.

<그림 2> 은 각 변수별 결측치 분포를 확인하는 bar graph이다. 이를 통해 어떤 변수에 어느 정도의 결측치가 존재하는지 패턴을 확인할 수 있다.

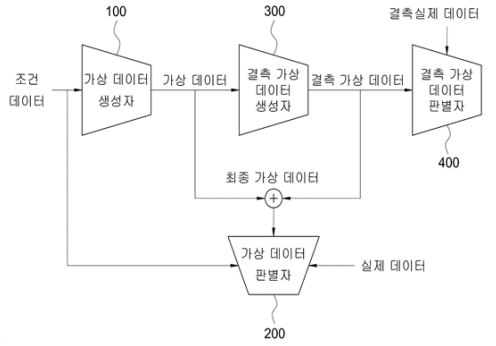


〈그림 2〉 결측치 패턴 확인

위에서 패턴을 파악한 후 결측치는 모델에 들어가는 형태로 맞춰주기 위해 label encoding 작업을 거치게 된다. 해당 과정에서는 Label information이라는 dictionary 형태의 빈 공간을 만들어 준 뒤, 결측치 label encoding을 한 값과 이에 해당하는 변수들의 정보를 담아둔다. 해당 정보들을 pickle 형태의 파일로 별도 저장을 진행한다.

2.4 GAN 모델 학습

날짜 및 공간 조건 데이터 및 결측치 데이터에 대한 전처리가 완료된 학습 데이터를 활용하여 GAN 모델을 학습한다. GAN 모델은 다음 도면의 과정으로 학습이 진행된다.



〈그림 3〉 GAN 모델 학습

GAN 모델 학습 시 임의의 분포를 가정하여 생성한 모형에서 학습된 잠재 변수에 날짜 및 공간 조건이 반영된 Deep-Convolution Layer를 사용한다. 다음으로 이전 layer의 모든 노드가 다음 layer의 모든 노드에 연결된 layer인 Full-connected layer도 함께 사용한다. 즉, GAN 모델 학습은 Deep-Convolution Layer와 Full-connected layer를 결합하여 가상 데이터를 생성하는 방식으로 이루어진다.

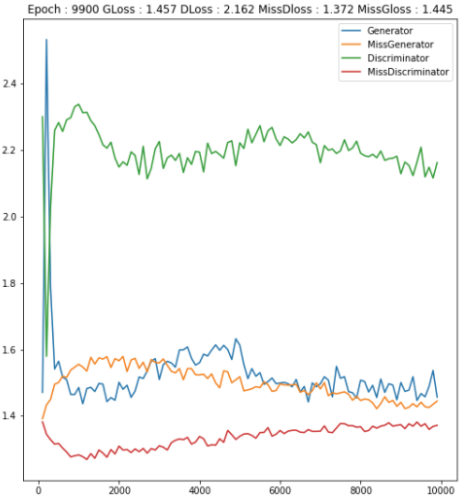
위 과정을 통해 생성된 가상 테이터에 결측 패턴을 반영하여 결측 가상 데이터를 생성한다. 이렇게 생성된 최종 가상 데이터를 GAN 네트워크의 Generator를 통해 입력을 받는다. 그 후 상기 날짜 및 위치 조건을 기반으로 생성된 최종 가상 데이터가 실제 데이터와 차이가 있는지 GAN 네트워크의 Discriminator를 통하여 판별한다.

최종적으로 학습된 데이터의 loss 값은 별도로 지정해둔 모델 학습 결과 경로에 학습이 진행될 때마다 쌓이면서 저장이 된다.

Ⅲ. 연구 결과

3.1 GAN 모델 성능 평가

학습된 GAN 모델을 이용하여 실제와 유사한 가상 데이터를 생성하기 이전에 모델의 성능을 파악하였다. 모델의 성능을 파악하기 위해 모델이 학습되면서 저장된 loss 값을 확인하였다. 아래 <그림 4>는 GAN 모델 학습을 통해 나온 loss 결과값 이다.



〈그림 4〉 Generator & Discriminator Error Graph

일반적인 네트워크 학습의 목적은 loss 값을 감소시키는 것이다. GAN 네트워크는 각 generator와 discriminator가 반대의 목적을 가지고 학습을 하기에 loss 값 또한 서로 반대의 의미를 가지고 있다. 그렇기에 generator와 discriminator의 loss 값이 균등하게 유지된다는 것은 어느 한쪽으로 편향되어 학습되는 것이 아닌 균형감 있게 학습이 되고 있음을 의미한다. <그림 4>을 보면, 학습이 진행될수록 특정 값 주변에서 각 generator와 discriminator의 loss 값이 잘 유지되는 것을 확인할 수 있다.

이를 통해 학습한 GAN이 어느 한쪽에 편향되지 않고, 균형감 있게 학습되어 전반적인 추세를 유지하고 있음을 확인하였다. 이를 통해 학습된 GAN을 가상 데이터 생성에 적용해보았다.

3.2 학습된 GAN 모델을 통한 최종 가상 데이터 생성

학습된 GAN 모델을 이용하여 실제 예측 모형에 사용될 수 있는 가상 데이터를 만들어 주기 위해 학습된 모델을 불러온다. 학습된 모델 이외에도 모델 학습을 위해 별도로 저장해둔 one-hot encoding, min-max scaling 등의 정보도 함께 불러온다.

이 과정에서는 학습된 GAN으로 가상 데이터를 생성하기 위해 특정 날짜 및 위치의 조건을 지정해준다. 여기서 지정해준 특정 발생 날짜 및 위치 조건을 바탕으로 학습된 GAN 모델이 해당 조건에 맞는 최종 가상 데이터를 생성한다. 해당 과정을 통해 생성된 최종 가상 데이터는 각 변수별로 값을 출력하며 동시에 결측값도 포함하여 값을 생성하기에 실제와 유사한 형태로 이루어져 있다. <그림 5> 은 GAN 시뮬레이션을 통하여 실제와 유사하게 생성된 가상 데이터의 형태이다.

Out[2]:

	chicken	turkey	duck	goose	quail	ornamental_bird	pheasant	ostrich	wild_bird	wild_duck	...	gate_U_shape_spray	ga
0	587767.750000	0.810081	34377.039062	0.407525	758194.750	0.611132	0.365650	0.271577	0.635084	0.738392	...	0.0	
1	485131.406250	0.204781	16623.117188	0.138387	502288.750	0.629470	0.275841	0.332693	0.370751	0.427687	...	0.0	
2	345493.437500	0.450622	9434.541016	0.605223	614331.500	0.302029	0.570665	0.306919	0.743683	0.259650	...	0.0	
3	436900.812500	0.727074	9196.843750	0.665399	541250.125	0.229799	0.576833	0.680720	0.281849	0.641084	...	NaN	
4	100044.890625	0.063515	18168.246094	0.610436	645496.000	0.848736	0.812091	0.761037	0.597715	0.597871	...	NaN	

5 rows × 45 columns

〈그림 5〉 최종 가상 데이터

3.3 연구 적용

본 연구에서는 GAN 학습을 통해 실제와 유사한 가상 데이터를 생성하였다. 생성된 가상 데이터는 실제 HPAI 확산 예측 모델의 input으로 들어갈 수 있는 데이터 마트의 형태로 변형하여 실제 데이터와 얼마나 유사하게 위험도를 감지하는지 확인하는데 사용될 수 있다는 점에서 의의가 있다.

해당 연구를 통해 생성된 가상 데이터를 확산 예측 모형에 들어가는 데이터 마트의 형태로 변형해준 뒤 HPAI 확산 예측 모델에 적용하여 실제와 유사한 성능을 보이는지 검증해보았다. 다음은 가상 데이터로 생성한 데이터 마트 일부와 예측 모델에 대한 결과 값이다.

farm_no	dt	chicken	duck	goose	ornamental_bird	ostrich	pheasant	quail	turkey	wild_bird	wild_duck
AAAA1	2019-01-01	587767	34377	0	0	0	0	758194	0	0	0
AAAA10	2019-01-01	480445	31277	0	0	0	0	683876	0	0	0
AAAA11	2019-01-01	436900	9196	0	0	0	0	541250	0	0	0
AAAA12	2019-01-01	295400	33569	0	0	0	0	886604	0	0	0
AAAA13	2019-01-01	100044	18168	0	0	0	0	645496	0	0	0
AAAA2	2019-01-01	485131	16623	0	0	0	0	502288	0	0	0
AAAA3	2019-01-01	769565	20064	0	0	0	0	242138	0	0	0
AAAA4	2019-01-01	432814	8340	0	0	0	0	748745	0	0	0

〈그림 6〉 GAN 학습을 통해 생성된 데이터 마트

RF1_P1	GBM1_P1	XGB1_P1	RF2_P1	GBM2_P1	XGB2_P1
0.001712058	0.000124928	1.43E-10	0.000893241	1.03E-05	2.11E-14
0.001712058	0.000102904	5.35E-11	0.000930988	9.18E-06	1.09E-14
0.001712058	0.000100471	6.94E-10	0.000858413	1.15E-05	2.00E-13
0.001712058	8.29E-05	7.07E-11	0.000794163	8.99E-06	2.38E-14
0.001712058	7.66E-05	5.34E-11	0.00069889	8.71E-06	3.43E-15
0.001712058	7.53E-05	3.08E-12	0.000666101	8.78E-06	1.05E-15
0.001712058	7.79E-05	2.00E-11	0.000739663	8.45E-06	2.81E-16
0.001712058	7.74E-05	5.76E-11	0.000818544	8.47E-06	5.44E-15
0.001759931	9.98E-05	5.59E-10	0.00083579	1.14E-05	4.97E-14
0.001765341	0.000137604	4.98E-09	0.001037345	1.69E-05	2.11E-12
0.001712058	0.000153884	8.36E-10	0.000954137	1.64E-05	8.34E-13
0.001712058	0.000115001	8.89E-10	0.001023624	1.08E-05	3.02E-13
0.001712058	0.000100471	7.48E-11	0.000801373	8.99E-06	9.04E-15
0.001712058	8.78E-05	1.95E-10	0.001157029	1.26E-05	1.10E-13
0.001712058	8.45E-05	9.62E-11	0.000901835	9.40E-06	5.54E-14
0.001717009	0.0001141	1.34E-09	0.001143657	1.48E-05	8.84E-13

〈그림 7〉 GAN으로 검증한 예측 모형의 위험도

GAN 학습으로 생성된 가상 데이터를 실제 확산 예측 모델에 적용하여 검증해보았을 때 실제 데이터로 기존 모델을 학습한 것과 유사한 성능을 보이는 결과를 확인하였다.

결론적으로 반복적인 GAN 학습을 통해 실제와 유사한 가상 데이터를 생성하면, 충분한 양의 학습 데이터를 확보할 수 있게 되기에 HPAI의 특성으로 인해 야기되는 충분하지 않은 데이터의 양과 데이터 내의 결측치 문제를 해결할 수 있다는 의의가 있다. 즉, 충분한 데이터양을 확보하게 된다면 기존 예측 모형의 성능을 저하시키는 원인 중 하나인 충분하지 않은 데이터양의 문제를 해결 및 보완하여 예측 모델의 성능을 높이는 것을 기대할 수 있다.

3.4 연구 한계 및 제언

3.4.1 가상 데이터의 한계

본 연구에서 GAN 학습을 통해 생성한 가상 데이터는 시뮬레이션을 통해 새롭게 생성된

sample data이다. 그렇기에 실제 HPAI 데이터와 비교하였을 때 얼마나 유사한지 측정할 수 있는 정량적인 척도가 존재하지 않는다. 해당 연구에서는 기존 예측 모델에 적용하여 성능을 통해 확인해 보았지만, 이 또한 모델 자체의 성능으로 인한 한계점을 가질 수 있다. 즉, 기존의 HPAI 확산 예측 모델의 성능에 따라 가상 데이터를 적용해보았을 때 성능이 달라지는 변동 사항이 생기기에 객관적인 판단을 할 수 없다.

3.4.2 연구 제언

해당 연구를 통해 생성된 가상 데이터를 연구에 실효성 있게 활용하기 위해서는 다음과 같은 과정이 필요하다.

먼저 실제 HPAI 확산 예측 모델에 사용되는 데이터와의 비교 검증이 필요하다. 비교 검증을 위해서는 GAN을 통해 생성된 가상 데이터의 변수와 실제 데이터의 변수의 다양한 지표들을 활용하여 가설 검정을 진행할 수 있다. 이를 통해 생성된 가상 데이터가 실제 데이터와 유사하게 생성이 되었는지 검증한 뒤 HPAI 확산 예측 모델의 성능을 향상을 위한 학습 데이터로 사용하여 실제 데이터로 학습하였을 때보다 성능이 향상되었는지 확인해볼 수 있을 것이다.

이에 추가적으로 충분한 양의 데이터로 확산 예측 모델에 직접 학습시켜봄으로써 실제 데이터와 비교하여 성능 향상에 실제로 데이터양이 미치는 영향력을 연구해볼 수 있다. 이에 충분한 데이터양이 성능에 미치는 영향력을 확인한다면, HPAI 확산 예측 모델 성능에 학습 데이터양 이외에 미치는 다른 원인들을 추가 발굴할 수 있으며 향후 더 많은 연구를 통해 HPAI 확산 예측 모델의 높은 성능 도출이 가능할 것이다.

참 고 문 헌

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

[2] Ian J.Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio.(2014) Generative Adversarial Nets.

[3] 최대우, 강태훈, 송유한 & 한예지. HPAI 다이나믹 데이터 마트와 설명 가능한 AI(XAI), J.Basic Sci., HUFS, 50(2020), p121-135.

사 사

이 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임.

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

한 예 지(Ye-Ji Han)

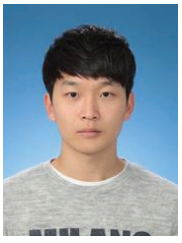
·2019년: 한양대학교 문화인류학과 (학사)
·2019년~현재: 한국외국어대학교 통계학과 (석사과정)
·관심분야: 빅데이터 분석, 자연어 처리, 텍스트 마이닝

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



이 원 빈(Won-Been Lee)

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