



J. Korean Soc. Aeronaut. Space Sci. 52(1), 77-86(2024)

DOI: <https://doi.org/10.5139/JKSAS.2024.52.1.77>

ISSN 1225-1348(print), 2287-6871(online)

가우시안 혼합 모델을 이용한 대한민국 공역 내 항공기 교통량 분석 및 예측

강진혁¹, 류재영², 이학태³Analysis and Prediction of Aircraft Counts in Korean National Airspace
Using Gaussian Mixture ModelJin-Hyeok Kang¹, Jaeyoung Ryu² and Hak-Tae Lee³

Dept. of Aerospace Engineering, Inha University, Incheon, Republic of Korea

ABSTRACT

For efficient air traffic management, it is essential to understand the traffic characteristics of the current operation and to be able to predict the traffic volume and capacity. In this paper, the change in aircraft count in all sectors and terminal maneuvering areas in the Incheon Flight Information Region is studied using one-year trajectory data. After obtaining the distribution of traffic volume change in one day, the maximum allowed aircraft was obtained for all airspaces. For further investigation, a machine learning-based clustering technique, Gaussian Mixture Model was used to cluster the change in aircraft count for each airspace. The results show that for each airspace, the daily change in aircraft count follows several patterns. Finally, it was shown that a certain level of prediction can be made to predict traffic changes and capacity from given traffic data by comparing it with the clusters.

초 록

효율적인 항공교통관리 체계를 구축하기 위해서는 각 공역에서 교통량의 현황을 파악하고, 데이터 분석을 기반으로 교통량 변화의 추이와 수용량을 예측할 필요가 있다. 본 논문에서는 2019년의 항적 데이터를 이용하여 1년간 인천 비행정보구역 내의 지역관제구역과 접근관제구역에 대하여 항공기 교통량을 산출하고 교통량의 분포를 분석하였다. 분석 결과를 바탕으로 각 공역 별 공역 내 최대 항공 교통량을 산정하였다. 좀 더 면밀한 분석을 위하여 머신러닝 기반 데이터 군집화 기법인 가우시안 혼합 모델을 활용하였다. 각 공역별 교통량의 변화를 군집화 하였으며, 하루 중 교통량의 변화가 몇 가지 패턴으로 나누어짐을 확인하였다. 최종적으로 학습된 군집과 대조하는 방식으로, 주어진 교통량 데이터로부터 교통량의 변화와 수용량을 예측하는 연구를 수행하여 어느 정도 예측이 가능함을 보여주었다.

Key Words : ATM(항공교통관리), Aircraft Count(공역 내 항공기 수), Airspace Capacity(공역 수량), Sector(지역관제구역), TMA(접근관제구역), Gaussian Mixture Model(가우시안 혼합 모델), Clustering(군집화)

† Received : December 2, 2022 Revised : November 28, 2023 Accepted : November 30, 2023

¹ Master, ² Ph.D Student, ³ Professor

³ Corresponding author, E-mail : haktae.lee@inha.ac.kr, ORCID 0000-0002-0674-9931

© 2024 The Korean Society for Aeronautical and Space Sciences

I. 서 론

COVID 19로 인해 전 세계 항공 교통량이 일시적으로 감소하였지만, 장기적으로는 여객 수요가 정상화되고 추가적으로 드론, 도심항공교통(UAM) 등과 같은 차세대 항공교통운송수단이 상용화되어 항공교통량은 꾸준히 증가할 전망이다[1]. 이로 인해 공역 내 교통은 지속적으로 더욱 혼잡해질 것으로 예상되며, 이를 관리할 항공교통관리 체계에 대한 다양한 연구가 진행되고 있다[2]. 그 중 항공교통량과 공역의 혼잡도간 상관관계를 분석하기 위한 연구가 집중적으로 수행되고 있다.

전 세계 대부분의 국가에서는 자국이 담당하는 항공관제구역을 지역관제구역(Sector)와 접근관제구역(Terminal Maneuvering Area, TMA)으로 나누어 보다 효율적으로 항공교통량을 관리하고 있다. 대한민국은 인천 비행정보구역(Flight Information Region, FIR)을 12개의 Sector와 14개의 TMA로 세분화하여 관리하고 있으며, Fig. 1에 각 공역의 3차원 형상을 나타내었다. 항공 교통량은 각 세부 공역별, 시간에 따라 달라지며, 특정 공역의 특정 시간대에 교통량은 관제사 업무부하의 한계에 근접할 수 있다. 이때의 교통량을 수용량이라 할 수 있으며, 이를 산정하고 예측하는 것은 항공교통흐름관리(Traffic Flow Management, TFM)와 같이 상위 단계에서 수요를 관리하여, 원활한 교통 흐름을 유지할 수 있게 해 주는 기법의 적용에 있어서 바탕이 된다.

Federal Aviation Administration은 특정 Sector의 교통량이 관제 시스템의 효율성에 영향을 미칠 수 있는 수준에 도달할 경우, 이를 경고하는 Monitor Alert Parameter (MAP)을 제시하였다[3]. MAP은 Sector를 지나가는 각 항공기당 관제사가 약 36초의 시간이 필요하다는 가정하에, 일정 시간동안 Sector를 지나가는 항공기 수가 기준값을 넘어갈 경우 alert값을 제시한다. 이는

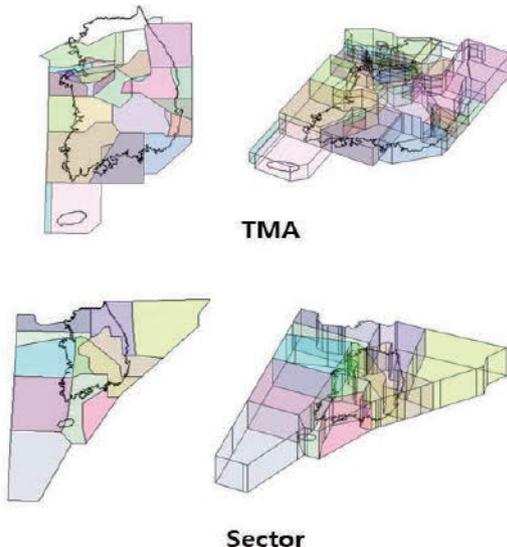


Fig. 1. Airspaces (TMA, Sector) in Incheon FIR

Sector가 가지는 교통 수용량의 간단한 모델이 될 수 있지만, 각 Sector의 실질적인 교통 수용량을 제시하지 못하는 한계를 가진다. Janic, M.은 항공기 착륙 시 사용하는 Terminal Gate, 항공기 타입 등에 정보를 파라미터화하여, Terminal Airspace의 수용량을 계산하는 방법을 제시하였다[4]. [4]의 경우는 항공기가 착륙하는 Terminal Airspace에서의 수용량을 계산할 수 있지만, En-route를 포함한 전체 공역의 수용량을 제시하기에는 한계가 있다. 또한, 최근 기계학습을 이용한 연구가 활발히 진행되면서 Han, S. H.는 지도학습 회귀알고리즘 중 앙상블 알고리즘을 활용하여 공항의 시간당 출발편 수와 도착편 수를 계산하는 연구를 진행하였다[5]. Wang, S.는 Vector Regression, Random Forest, Artificial Neural Network 등의 기법을 이용하여 날씨에 따른 Terminal Airspace의 수용량을 제시하는 연구를 진행하였다[6].

본 논문에서는 2019년 ADS-B 데이터를 이용하여, 인천 FIR 내의 각 Sector와 TMA에서의 교통량을 시간대별로 산출하고, 공역 별 최대 항공 교통량을 제시한다. 기본적인 교통 수용량은 각 공역에서 관제사가 동시에 관제할 수 있는 항공기 대수로 정의하였다. 국내 공역을 통과한 항공기의 데이터를 비교하면, 2019년 국내 공역을 통과한 항공기가 2018년의 경우에 비해 약 2.2퍼센트 늘어난 것을 확인하였다. 이 데이터만으로 항공편의 수요가 충분하다고 판단할 수는 없지만, 항공편의 충분한 수요가 존재하였다고 가정하여, 최대 항공 교통량을 기본적인 교통 수용량으로 판단하였다. 또한 머신러닝 기반 데이터 군집화 기법인 가우시안 혼합 모델(Gaussian Mixture Model, GMM)을 이용하여 시간에 따른 각 공역별 교통량 변화의 패턴을 제시하고, 관제구역에서의 교통량을 예측한다.

본 논문의 구성은 다음과 같다. II 장은 항적 데이터를 통해, 각각의 Sector와 TMA에서 항공기 교통량을 산출하는 과정을 보여준다. III 장은 각 공역의 항공기 교통량 산출 결과를 토대로 군집화를 진행하여 데이터를 분류하고, 이를 통해 공역 교통량을 예측하는 과정 및 그 결과를 제시한다. IV장은 본 논문의 결론과 향후 연구 계획을 제시한다.

II. 공역 교통량 산출

2.1 대한민국 공역 좌표 전처리

각 공역의 3차원 형상과 항적 데이터를 비교하여 공역 교통량을 산출하기 위해 Aeronautical Information Publication (AIP)에 기재된 공역 좌표를 평면상에 투영할 필요가 있다. 본 연구에서는 람베르트 정각원추도법을 사용하여 Table 1의 ϕ_0 , λ_0 를 원점으로 하는 평면상에 투영하였다[7]. 좌표 투영 시 사용한 기준값은 Table 1과 같다.

Table 1. Reference Coordinates for Lambert Conformal Projection

Reference	Value	Unit	Reference	Value	Unit
ϕ_{min}	30.0	deg	ϕ_{max}	38.7	deg
λ_{min}	124.0		λ_{max}	133.7	
ϕ_0	30.0		λ_0	128.85	

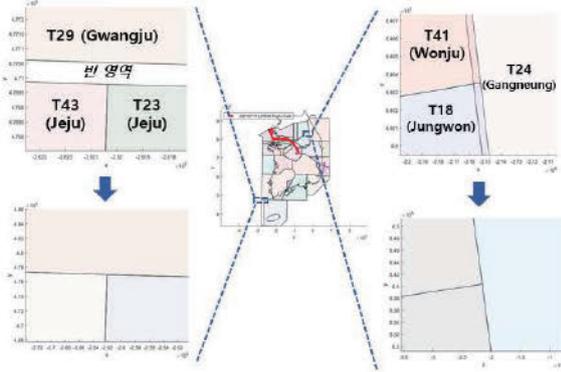


Fig. 2. Results of Airspace Coordinate Preprocessing [8]

람베르트 정각원추도법을 통해 투영된 Sector와 TMA의 좌표를 분석한 결과, 2개 이상의 공역들이 서로 맞닿은 영역에서 좌표 불일치로 인해 중첩되거나 빈 영역이 생성되었다. 관제와 같은 일반 운용 상황에서는 엄밀한 좌표의 일치가 필요하지 않으나, 수치적인 데이터 분석을 위해서는 빈 영역 또는 중첩을 제거한 보정된 좌표가 필요하다. 본 논문에서는 [8]에서 제시한 보정된 좌표를 사용하였으며, 그 중 일부를 Fig. 2에 나타내었다.

2.2 각 공역별 교통량 산출

2.2.1 ADS-B 항적 데이터

보정된 Sector와 TMA 형상과 항공기의 항적데이터를 이용하여 각 항공편 별로 특정 공역에 대한 진입과 진출 시각을 구하고, 이를 바탕으로 각 공역 내의 항공기 대수를 도출하였다. 이때 시간을 10분 간격으로 나누었으며, 10분 동안 공역 내에 존재한 항공기 최대 대수를 교통량으로 정의하였다.

본 연구에서는 항적데이터로 2019년 인천 FIR을 통과한 항공기 약 93만 대의 ADS-B 데이터를 이용하였다. Fig. 3에 단일 항공기에 대한 ADS-B 데이터를 나타내었으며, 항공기의 출발공항부터 도착공항까지의 항적 데이터와 운항 정보를 포함하고 있다.

Callsign	AAR108					
Type	A321					
Origin	RKSI					
Destination	RJAA					
Time	Lat (deg)	Lon (deg)	Alt (ft)	GSD (Knot)	VRT (fpm)	Hdg (deg)
2019-06-02 7:03:22 PM	37.49192	126.42738	1500	177	-	327
2019-06-02 7:03:39 PM	37.50448	126.41735	2250	179	2338	328
2019-06-02 7:03:56 PM	37.51598	126.40833	2825	182	2091	328

Fig. 3. ADS-B Trajectory Data Format

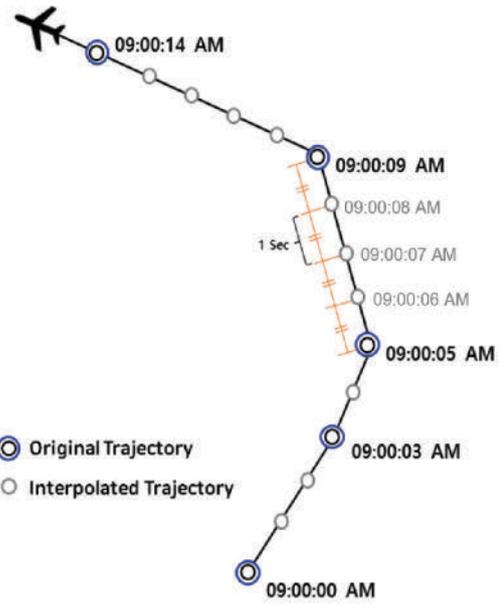


Fig. 4. ADS-B Data Interpolation

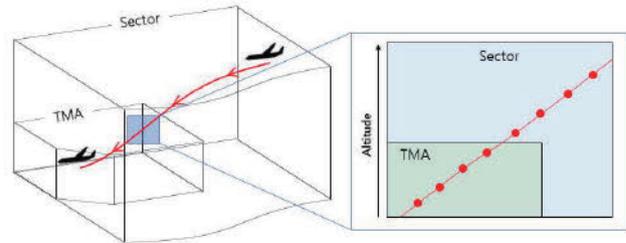


Fig. 5. Result of Passed Airspace Change.

2.2.2 공역 내 항적 진입 여부 판별

본 논문에서 사용한 ADS-B 데이터의 경우, 데이터 기록 주기가 일정하지 않기 때문에 시간동기화가 필요하다. Fig. 4와 같이 항적 데이터를 1초 단위로 선형 보간하여 동기화를 진행하였다[9].

보간된 항적 데이터의 각 점이 어느 공역 내부에 속해 있는지를 판별하기 위해, Point in Polygon 알고리즘 [10]을 사용하여 항적 좌표와 공역 좌표의 2차원 데이터를 비교하였다. 위의 과정을 통해 2차원 상에서 항적 좌표가 공역 내부에 존재한다고 판명되었을 경우, 항적 좌표의 고도와 공역의 최대·최소 고도범위를 비교하여, 항공기가 3차원 공간 내부에 존재하는지 판별하였다. 항공기가 현재 공역에서 다음 공역으로 이동할 때, 공역 간의 고도 범위 내 수평면 상에서 경계를 넘어 이동하는 경우와 Fig. 5와 같이 고도 경계면을 지나는 경우가 있음을 확인하였다.

2.2.3 2019년 각 공역의 교통량 산출

1초 간격으로 선형 보간된 모든 항적 좌표에 해당하는 공역을 계산한 후에 각 공역의 시간에 따른 항공기 대수의 증감을 계산하였다. 실제로 매 초마다 대수가 달라지지만, 이런 경우 대수의 변화가 너무 잦고 데이터 표

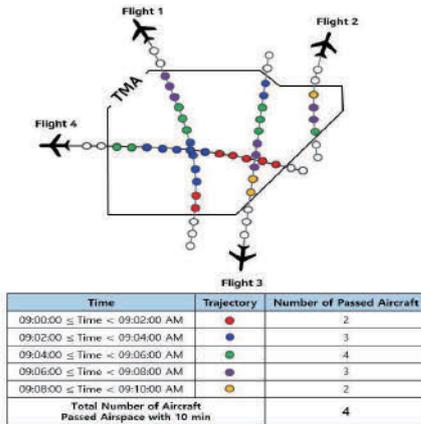


Fig. 6. Max Number of Aircraft Flying Simultaneously at Airspace with 10 minutes

현에 한계가 있어서 10분 간 최대 대수를 그 시간대의 교통량이라고 정의하였다. 시간 차원이 겹치는 부분이 없도록 rolling time을 10분으로 설정하였으며, 10분 간격으로 측정을 진행하였다. 시간 기준은 KST를 이용하여 연구를 진행하였다. Fig. 6은 4대의 항공기가 각각 시차를 두고 해당 공역에 진입했다가 진출하는 예시를 보여준다. 10분 동안, 항공기의 대수는 최소 2대에서 최대 4대까지 변하지만 최대값인 4가 되는 것을 보여준다.

산출된 각 공역의 항공기 교통량 결과 중 인천 북부 Sector와 제주 TMA의 결과를 Fig. 7과 8에 나타내었다. Fig. 7과 8의 상단 그림은 1년 전체에 대한 변화를 보여 주고 있으며, 하단 그림은 그 중 특정 하루에 대하여 0시부터 24시까지의 변화를 보여준다.

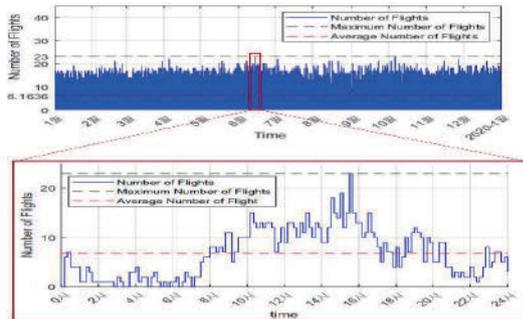


Fig. 7. Result of 2019 Number of Flights Passed Incheon North Sector

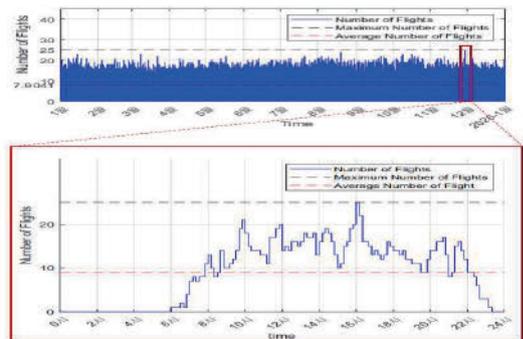


Fig. 8. Result of 2019 Number of Flights Passed Jeju TMA

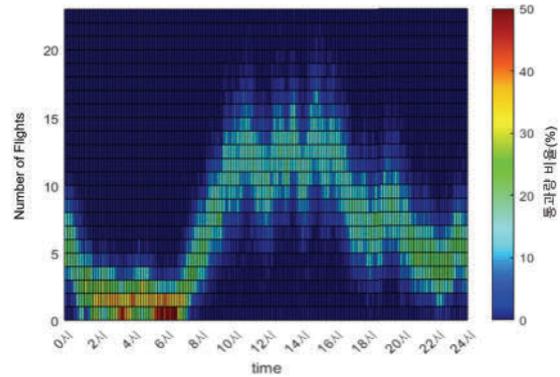


Fig. 9. Result about the Weight of Aircraft Count Passed Incheon North Sector in 2019

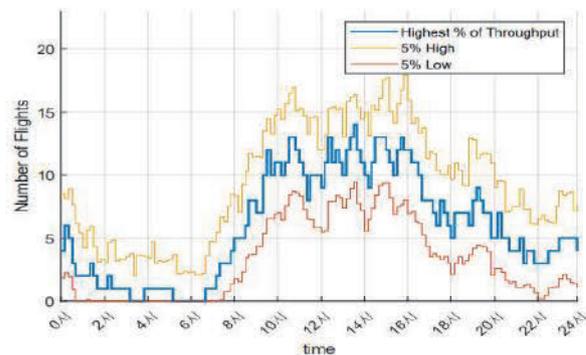


Fig. 10. Top 5%, Median, and Bottom 5% percentile lines for Aircraft Count in Incheon North Sector in 2019

III. 공역 수용량 산정 및 교통량 예측

3.1 각 공역의 1일 교통량 비교분석

대부분의 항공편이 정기운항편이기 때문에, 일자별 시간대에 따른 교통량의 변화 패턴은 크게 달라지지 않는다고 가정하여, 각 공역별로 1년간의 교통량을 하루 24시간에 중첩하여 분포를 분석하였다. Fig. 9는 인천 북부 Sector에 대한 교통량 분포를 보여주며 열은 녹색으로 나타나는 부분이 가장 혼한 교통량 변화 패턴이다.

이와 같이 생성된 분포를 이용하여 최빈값, 하위 5%, 상위 5% 값을 산출하였다. Fig. 9에 나타난 연속적인 분포에 대해 하위 5%, 최빈값, 상위 5% 선을 Fig. 10에 나타내었다. 위의 방식으로 교통량을 분석한 결과, 대부분의 공역에서 새벽과 저녁 시간대에 교통량이 급감하는 것을 확인하였다. 한편 높은 수준의 교통량이 유지되는 10시부터 16시까지는 대부분의 공역에서 교통량의 상한이 유지되는 특성을 발견할 수 있었다. 실제로 이러한 상한이 항공편의 수요의 한계에서 오는 수용량의 상한인지, 아니면 관제사 업무 부하의 한계에서 오는 수용량의 상한인지 명확히 구분하기는 어렵다. 하지만 국내 공역을 통과한 항공편의 수가 2018년도에 비해 2.2퍼센트만 증가하였기 때문에 항공편의 충분한 수요가 존재하였다고 가정하여, 해당 지표를 기본적인 수용량이라고 판단하였다.

Table 2. Results of Maximum Aircraft Count from the Top 5% Percentile and Corresponding Time Period for Sectors

Sector	Max Number of Aircraft	Time
Gwangju East	26	05:50 ~ 06:00
Pohang Area	18	10:20 ~ 10:30
East-Sea	12	10:50 ~ 11:00
Gangneung	14	10:50 ~ 11:00
Incheon South	25	15:20 ~ 15:30
Incheon North	18	15:50 ~ 16:00
Jeju Area	35	15:50 ~ 16:00
Gunsan East	17	16:40 ~ 16:50
South Area	11	18:00 ~ 18:10
Gunsan West	17	19:30 ~ 19:40
Daegu Area	14	20:20 ~ 20:30
Gwangju West	20	20:50 ~ 21:00

Table 3. Results of Maximum Aircraft Count from the Top 5% Percentile and Corresponding Time Period for TMAs

TMA	Max Number of Aircraft	Time			
Sacheon	9	05:40 ~ 05:500			
Gimhae	15	09:10 ~ 09:20			
Seoul	43	10:50 ~ 11:00			
Jeju	20	14:20 ~ 14:30			
Gwangju	14	16:40 ~ 16:50			
Osan	9	18:10 ~ 18:20			
Daegu	11	20:10 ~ 20:20			
Haemi	Pohang	Gunsan	Yecheon	Wonju	Gangneung
Max Number of Aircraft Passed Airspace is 4 or less.					

1일 기준 교통량 분포에서 상위 5% 비중을 가지는 값 중 최대 교통량을 가지는 시간대와, 이때의 교통량의 값을 Sector와 TMA에 대하여 각각 Table 2와 3에 나타내었다. 본 논문에서는 Table 2와 3에 제시된 최대 교통량 값을 각 공역별 기본적인 수용량으로 판단하며, 제시한 시간대가 각 공역별로 가장 혼잡한 시간대라고 판단한다. 최대 교통량이 4대 이하인 공역은 결과에서 제외하였다.

3.2 공역 교통량 데이터 군집화

3.2.1 각 공역별 1일 교통량 산출 결과 분석

이전 절에서 각 공역별 1일 교통량을 분석한 결과, 대

다수의 항공기가 일정 스케줄에 따라 정해진 항로로 비행하는 대한민국 공역 내 항공 운항 특성상, 각 공역별 시간대에 따른 교통량 증감의 추이가 비교적 일정함을 확인하였다. 이에 착안해 대표적인 교통량 증감 패턴을 확인할 수 있다면, 특정 시간대의 항공 교통량을 예측할 수 있을 것이라 판단하였다. 이를 위해 세부적으로 교통량 증감 패턴이 유사한 일자별로 교통량 데이터를 분류하는 기법이 필요하였고, 공역의 전체 항공기 교통량 데이터를 특정 개수의 확률 분포로 나타낼 수 있는 가우시안 혼합 모델(GMM)[11]을 사용하여 교통량 군집화(Clustering)를 진행하였다.

3.2.2 가우시안 혼합 모델

가우시안 혼합 모델은 분석 대상 데이터들이 가우시안 분포를 따를 것이라는 가정에서부터 시작된다. 모든 데이터가 하나의 가우시안 분포를 따르기에는 데이터의 경향성이 일정하지 않기 때문에, 다수의 가우시안이 중첩된 분포를 따른다고 가정하여 각각의 가우시안 분포들을 찾는 것을 목적으로 한다.

각각의 가우시안 분포는 평균값을 의미하는 μ_k , 분산 값을 의미하는 Σ_k , 전체 분석 대상 데이터에서 해당 가우시안 분포의 영향력을 의미하는 혼합 확률 π_k 로 구성된다. 가우시안 혼합 모델은 모든 데이터를 기준으로 하여 가장 높은 GMM possibility를 찾는 것을 목적으로 하며, 이를 찾기 위한 가우시안 확률 밀도 함수를 식 (1)에 나타내었다.

$$p(x) = \sum_{k=1}^K \pi_k N(x|\mu_k, \Sigma_k) \tag{1}$$

기댓값-최대화(Expectation-Maximization) 알고리즘을 이용하여 GMM Possibility를 가장 높게 만드는 μ_k, Σ_k, π_k 를 찾게 된다. Expectation 단계에서는 모든 데이터에 한하여 각각의 가우시안 분포에 속할 확률(Responsibility)를 구하게 된다. 이를 이용하여 log-likelihood function을 구할 수 있으며 이는 식 (2)와 같다.

$$\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k N(x_n|\mu_k, \Sigma_k) \right\} \tag{2}$$

Maximization 단계에서는 log-likelihood 값을 최대화 하기 위한 μ, Σ, π 를 계산하게 되며, 이는 주어진 모든 데이터에 대해 가장 높은 GMM possibility를 찾는 것을 의미한다. Expectation과 Maximization은 μ, Σ, π 가 수렴하거나 일정 횟수를 반복할 때까지 진행되며, 각각의 데이터는 Responsibility가 가장 높은 가우시안 분포에 속하게 된다.

가우시안 혼합 모델은 k개의 군집이 존재한다고 가정하기 때문에 사용자가 직접 k값을 지정해 줄 필요가 있다. 본 연구에서 k값은 계층적 군집화 방식[12]을 이용해 계산하였다. 계층적 군집화 방식 알고리즘을 실행하

기 위해, 각 군집의 중심값 사이의 문턱거리를 구하는 과정에서 노이즈나 이상치에 덜 민감한 장점이 있는 ward 연결법[13]을 이용하였다. ward 연결법의 문턱거리를 구하는 식은 식 (3)에 나타내었다.

$$d(A, B) = \frac{\|\mu_A - \mu_B\|^2}{\frac{1}{n_A} + \frac{1}{n_B}} \quad (3)$$

식 (3)에서 μ_A 와 μ_B 는 군집 A와 B의 평균값이며, n_A 와 n_B 는 군집 A와 B의 데이터 수이다. 식 (3)을 이용하여 도출한 대구 TMA의 cluster 결과를 가시화한 dendrogram은 Fig. 11과 같다. 클러스터별 중심값 사이의 거리를 의미하는 문턱거리가 커짐에 따라 군집의 숫자가 작아짐을 볼 수 있는데, 문턱거리가 70 근방인 지점에서 문턱거리가 달라져도 군집의 갯수가 크게 변하지 않는 것을 확인하였다. 본 연구에서는 하루를 십분 단위로 나누어 총 144개의 차원을 분석하였기에, 각 차원별 문턱거리를 0.5로 설정하여 문턱거리를 72로 설정하였다. 이는 클러스터 간 비교를 할 때, 각 시간 차원에 제시된 교통량 값의 차이가 평균적으로 0.5대를 넘지 않는 것을 의미하며, 평균적으로 0.5보다 커질 경우, 다른 클러스터로 판별되는 것을 의미한다. 한편 클러스터의 수가 많아지는 공역의 경우, 차원별 문턱거리를 1로 설정하여 문턱거리를 144로 설정하였다. 서울 TMA와 제주 Sector의 경우는 문턱거리를 72로 할 경우 군집의 숫자가 많아져서 문턱거리를 144로 높여 연구를 진행하였다.

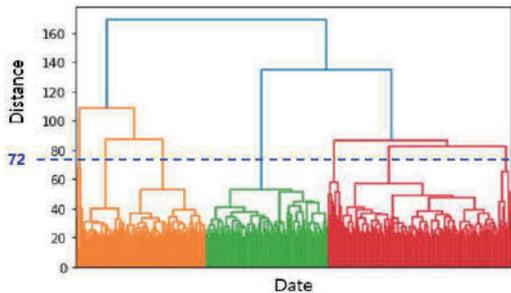


Fig. 11. Dendrogram: Number of Aircraft Passed Daegu Airspace

Table 4. Number of Cluster in each Airspace: k

Airspace		Number of Cluster: k
Type	Name	
TMA	Seoul	8
	Jeju	9
	Gwangju	8
	Daegu	3
Sector	Incheon South	11
	Jeju Area	7
	Pohang Area	8
	East-sea	3
	Daegu Area	7

해당 과정을 통해, 각 공역별 항공기 교통량 데이터 군집화에 필요한 최적의 군집 개수 k를 산출하였으며, 이를 Table 4에 정리하였다.

3.2.3 공역 교통량 데이터 군집화

본 논문에서 사용한 데이터는 하루를 기준으로 144개의 차원을 가지며, 교통량 예측을 위해 제외한 12월 결과를 제외하고 11개월의 데이터를 이용하여, 총 334일에 해당하는 데이터로 연구를 진행하였다. 각 공역의 교통량은 시간대별로 실제로 독립적이지 않지만, 각 시간대별 영향력을 최소화하기 위해 대각 공분산(Diagonal Covariance)을 가지도록 군집화를 진행하였다.

3.2.4 가우시안 분포별 이상 탐지

이상 탐지[14]는 데이터들 중에 비정상적인 데이터를 찾는 알고리즘이다. 데이터별 각각의 군집의 속할 확률을 구한 다음, 특정 경계값보다 작을 경우 데이터의 이상이 있다고 판단한다. 데이터별 확률은 확률 밀도 함수를 사용하여 계산할 수 있으며, feature가 서로 독립일 경우, 각각 feature의 확률 밀도 함수의 곱으로 데이터의 확률 밀도 함수를 나타낼 수 있다. 공역 교통량은 시간별로 독립적이라 가정하였기 때문에, 데이터의 확률 밀도 함수는 시간 대별 확률 밀도 함수의 곱으로 나타낼 수 있으며, 데이터의 확률 밀도 함수는 식 (4)와 같이 표현할 수 있다. 여기서 m 은 시간 feature를 의미한다.

$$p(x) = p(x_1; \mu_1, \sigma_1^2) \dots p(x_m; \mu_m, \sigma_m^2) = \prod_{m=1}^M p(x_m; \mu_m, \sigma_m^2) \quad (4)$$

앞에서 구한 가우시안 혼합 모델은 대각 공분산 행렬을 가지고 있으며, 공분산 행렬의 $m \times m$ 행렬값은 각 feature의 σ_m^2 값을 의미한다. μ_m 은 혼합 모델 평균의 m 번째 값을 의미하며 이를 통하여 클러스터별 이상치를 짐작하거나 예측할 수 있다.

본 연구에서는 이상 탐지를 위해 클러스터별 경계값을 구성하였으며, 경계값은 각 feature별 $\mu \pm \sigma$ 값을 이용하여 구성하였다. 정규분포에서 $\mu \pm \sigma$ 의 데이터가 가지는 확률값은 약 0.24이기 때문에, m 번째 값을 제외하고 다른 값들이 평균값 μ_k 를 따르고, m 번째 값이 $\mu \pm \sigma$ 의 값을 갖는 데이터의 경우, 약 0.24의 Responsibility를 가지게 된다. 클러스터별 경계값을 이용하여 공역 수용량을 추정하였다.

3.2.5 공역 교통량 데이터 군집화 결과

본 논문에서는 공역 교통량이 타 공역보다 상대적으로 많은 서울, 제주, 대구, 오산 TMA와 인천 FIR 내 모든 Sector에 대해 교통량 데이터 군집화를 진행하였다. 이 중 교통량이 타 공역 대비 상대적으로 많은 서울 TMA와 제주 Sector의 결과를 살펴보았다. 서울 TMA와 제주 Sector의 클러스터별 교통량 일 수를 Table 5에 정리하였다.

Figure 12와 13은 서울 TMA 교통량 군집화 결과 중 7번 군집과 0번 군집의 모습을 나타내고 있다. 해당 그래프의 검은 선은 군집의 중심값(Centroid)을 나타낸 것

이며 빨간 선은 중심값으로부터 $+\sigma$ 범위를, 파란 선은 중심값으로부터 $-\sigma$ 범위를 나타낸다. 해당 공역의 군집화 결과를 비교하였을 때, 0번 군집을 제외하고는 교통량의 증감이 모두 비슷한 경향을 가지고 있는 것을 확인하였다. 0번 군집의 경우, 해당 군집의 속하는 데이터 수가 다른 군집에 비해 적었으며, 데이터와 0번 군집의 중심값을 비교하였을 때에도 차이가 많이 나기 때문에 noise로 판단하였다.

Noise로 판단한 0번 군집을 살펴보면 총 4일의 데이터가 포함된다. 이 중 하나인 9월 7일의 데이터를 살펴보면 많은 수의 교통량이 유지되어야 하는 오후 두시부터의 교통량이 급격히 떨어진 것을 확인할 수 있었으며, 해당 군집의 다른 데이터인 11월 8일의 데이터 또한 오후 세시부터의 교통량이 급격히 떨어진 것을 확인하였다.

Table 5. The Number of Cluster Data

cluster	Seoul TMA	Jeju Sector
cluster 0	4	25
cluster 1	21	98
cluster 2	30	6
cluster 3	67	36
cluster 4	31	70
cluster 5	97	91
cluster 6	45	9
cluster 7	39	-

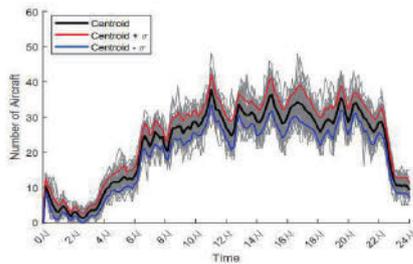


Fig. 12. Seoul TMA: Cluster 7

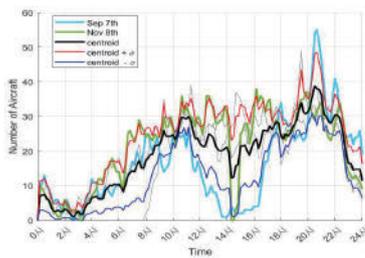


Fig. 13. Seoul TMA: Cluster 0

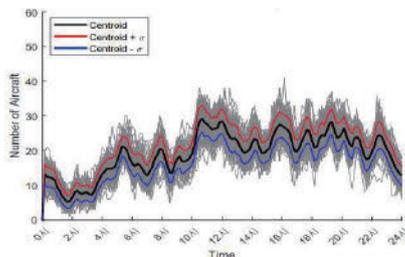


Fig. 14. Jeju Area Sector: Cluster 5

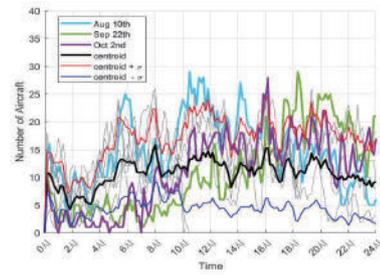


Fig. 15. Jeju Area Sector: Cluster 6

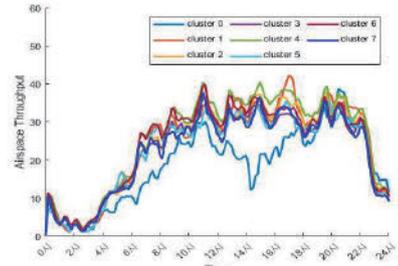


Fig. 16. Seoul TMA Cluster Centroid Set

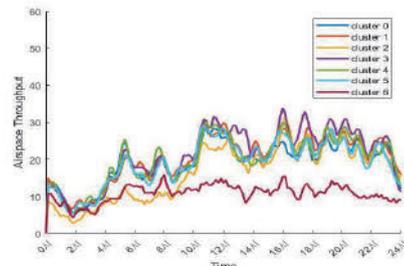


Fig. 17. Jeju Area Sector Cluster Centroid Set

Figure 14과 15의 경우 제주 Sector 교통량 데이터 군집 중, 5번 군집과 6번 군집을 나타낸 것이다. 제주 Sector는 6번 군집을 제외하고는 모두 비슷한 증감의 경향을 보이고 있었고, 6번 군집의 포함된 데이터 수가 다른 군집에 비해 상당히 적었기 때문에 해당 군집 데이터 또한 noise 결과로 판단하였다. noise로 분석된 6번 군집을 살펴보면 총 9일의 데이터가 포함되어 있음을 확인하였다. 해당 군집에 속하는 8월 10일, 9월 22일, 10월 2일 등의 데이터를 살펴보면, 6번 군집의 중심값과의 차이가 큰 것을 볼 수 있다.

Figure 16과 17의 각각 서울 TMA와 제주 Sector 각 군집의 중심값을 나타낸 것이다. 중심값은 noise 처리된 하나의 군집을 제외하고 모두 비슷한 증감의 경향을 나타내고 있으나 국소적으로 어느 정도 차이를 보이고 있음을 확인하였다.

3.2.6 공역 교통량 예측 결과

2019년 1월부터 11월까지의 교통량으로 구성된 가우시안 혼합 모델을 이용하여 2019년 12월 각 공역별 항공기 교통량을 예측하였다. 교통량 예측을 위해 구성한 가우시안 혼합 모델은 2019년도 데이터를 이용하여 해당 년도에 국한된 예측이 가능하다는 한계가 존재한다.

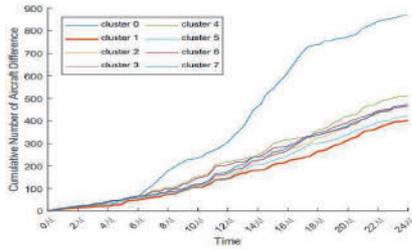


Fig. 18. Cumulative Errors of Cluster Centroids for the Aircraft Count of Seoul TMA in 2019.12.15

하지만 본 연구에서 제시한 방법을 바탕으로, 예측을 원하는 시점 기준으로 1년 간의 데이터를 이용하여 새로운 모델 구성이 가능할 것이다. 2024년 4월달의 예측 모델을 구성할 경우, 2023년 3월달의 데이터부터 2024년 3월달의 데이터를 이용하여 구성이 가능할 것이다.

2019년 12월 특정 구역과 날짜의 교통량을 해당 구역의 각 군집의 중심값과 대조하여 시간대 별로 오차를 계산하고 누적 오차가 가장 작은 군집의 중심값을 예측치로 산정하였다. Fig. 18은 서울 TMA의 2019년 12월 15일 교통량 데이터가 해당 구역의 8개의 군집의 중심값에 대하여 0시부터 계산된 누적 오차를 보여주고 있다.

교통량이 작은 6시까지는 군집별로 차이가 크지 않으나, 6시 이후의 데이터가 포함되기 시작하면 noise로 판단되었던 군집 0에 대한 누적오차는 급격하게 늘어나며, 나머지 군집들과의 차이도 조금씩 커지는 것을 볼 수 있다. 군집 1은 처음부터 일관되게 가장 작은 오차를 나타내기 때문에 2019년 12월 15일의 교통량 패턴은 군집 1에 해당함을 알 수 있다. 실제로 군집 1의 중심값과 실제 교통량을 비교한 것을 Fig. 19에 나타내었으며, 유사한 것을 볼 수 있다.

Figure 20은 제주 Sector의 2019년 12월 15일 교통량 데이터가 해당 구역의 7개의 군집의 중심값에 대하여 0시부터 계산된 누적 오차를 보여준다. noise로 판단된 군집 6의 경우, 초반에 오차가 적지만 시간이 지날수록 오차가 커지는 것을 볼 수 있다. 군집 3의 경우, 초반에는 오차가 커지지만 결국 noise로 판단된 군집 6보다 오차가 작아지는 것을 볼 수 있으며, 나머지 군집 중에 군집 0이 일관적으로 작은 오차를 가진다. Fig. 21은 제주 Sector에서의 12월 15일의 실제 교통량의 변화와 군집 0의 유사도를 제시한다.

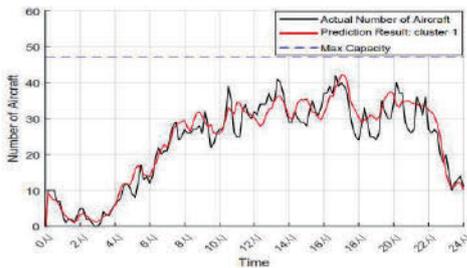


Fig. 19. Comparison between the Aircraft Count of Seoul TMA in 2019.12.15, and the Centroid of the Cluster with the Smallest Cumulative Error

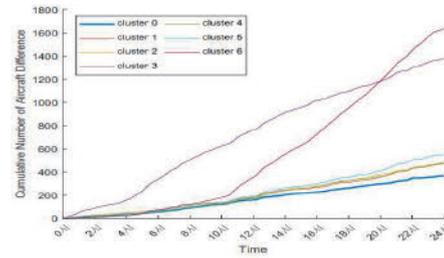


Fig. 20. Cumulative Errors for All Cluster Centroids for the Aircraft Count of Jeju Area Sector in 2019.12.15

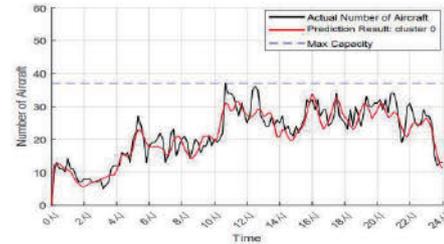


Fig. 21. Comparison between the Aircraft Count of Jeju Area Sector in 2019.12.15, with the Smallest Cumulative Error

3.2.7 구역 수용량 예측 결과

해당 날짜의 교통량 추이가 속하는 군집의 분포를 이용하여, 앞 절에서 상위 5% 교통량 중 최대 값을 수용량으로 산정한 것과 유사하게, $\mu + \sigma$ 에 해당하는 교통량의 최대값을 예측된 수용량으로 산정하였다. Fig. 18과 20을 보면 오전 10시 이전까지의 클러스터의 중심값과 교통량의 차이만으로는 어느 클러스터의 판별되는지를 파악하기 힘들지만, 이후부터는 어느 정도의 예측이 가능한 것을 볼 수 있으며 교통량이 최대가 되기 이전에 수용량 예측이 가능하게 된다.

GMM 모델의 학습에 사용하지 않은 12월의 데이터 중 7일에 대하여 수용량을 예측을 시도하였다. Fig. 22는 제주 Sector에서 0번 군집으로 판단된 테스트 항목들과 중심값을 나타낸 것이다. 테스트 궤적들과 0번 군집의 오전 10시 까지의 데이터를 비교하면, 교통량의 차이가 크지 않으며, 0번 군집으로 예측이 가능한 것을 보여준다. Table 6에 서울과 제주 TMA에서의 $\mu + \sigma$ 기준 수용량 예측 결과와 결과에 대한 평균, 분산값을 나타내었고, Table 7에는 인천과 제주 Sector에 대한 수용량 예측 결과를 나타내었다.

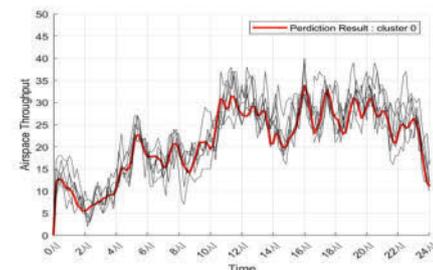


Fig. 22. Comparison between the Aircraft Count of Jeju Area Sector with Cluster 3 Centroid

Table 6. Capacity Prediction: TMA

Day	Cluster (Seoul)	Capacity $\mu + \sigma$	Mean / Standard vairation	Cluster (Jeju)	Capacity $\mu + \sigma$	Mean / Standard vairation
01	5	40	2.50 (3.37)	2	18	1.36 (2.03)
05	5		2.33 (2.98)	1		1.25 (1.89)
10	1	47	5.15 (6.60)	1		1.39 (2.05)
15	1		2.77 (3.61)	3		1.13 (1.60)
20	5	40	2.43 (3.20)	2		1.30 (1.86)
25	5		2.96 (3.81)	2		1.00 (1.43)
30	5		2.91 (3.77)	0		1.22 (1.74)

Table 7. Capacity Prediction: Sector

Day	Cluster (Incheon North)	Capacity $\mu + \sigma$	Mean / Standard vairation	Cluster (Jeju Area)	Capacity $\mu + \sigma$	Mean / Standard vairation
01	1	18	2.08 (2.68)	0	37	2.36 (3.09)
05			1.52 (2.02)			3.11 (4.06)
10			1.93 (2.55)			3.26 (4.12)
15			1.71 (2.27)			2.56 (3.21)
20			2.01 (2.73)			3.20 (4.02)
25			1.97 (2.55)			2.53 (3.28)
30			2.14 (2.97)			3.53 (4.38)

IV. 결 론

본 연구는 대한민국 관제 공역의 항공 교통을 분석하여 각 Sector와 TMA 공역별 교통량의 하루 중 변화 추이와 분포를 살펴보고, 이를 이용하여 공역별 최대 교통량을 산정하였다. 또한 머신러닝 기반의 데이터 군집화 모델인 GMM 모델을 이용하여 공역별로 교통량 데이터를 군집화하고, 이를 통해 각 공역별 교통량 변화의 패턴이 몇 가지로 분류될 수 있음을 확인하였다. 마지막으로 학습에 이용되지 않은 날짜에 대하여 데이터 기반으로 교통량과 수용량을 예측하는 것이 어느 정도 가능함을 보여 주었다.

본 논문을 통해 수행한 연구 결과를 통해 공역별 교통량의 현황을 파악할 수 있고, 생성된 교통량 결과는 안전 관리 및 교통 관리의 바탕이 될 수 있을 것이다. 특히 수용량 예측은 향후 흐름관리와 같이 스케줄링 관련된 연구에 유용하게 사용될 것으로 기대된다.

후 기

본 논문은 2022년도 정부(국토교통부)의 재원으로 '데이터기반 항공교통관리 기술개발(과제번호 : 22DATM-C1

63373-02)' 연구의 지원과 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.RS-2022000155915, 인공지능융합혁신인재양성(인하대학교))을 받아 수행된 연구로서, 관계 부처에 감사드립니다.

References

- 1) Kim, H. Y. and Yoo, K. E., "A Study on Aviation Industry of APAC Region and A Way of Safety Improvement," *Journal of the Aviation Management Society of Korea*, Vol. 16, No. 6, pp. 157~169.
- 2) Kim, Y. K., Won, I. S., Yun, J. C. and Park, H. D., "A Study On Automatic Control Sector Hand-off Algorithm For Air Traffic Control System Automation," *Journal of Advanced Navigation Technology*, Vol. 14, No. 4, pp. 489~494.
- 3) FAA Order 7210.3AA, Facility Operation and Administration, CHG 2, 13 Sep 2018
- 4) Janic, M. and Tosic, V., "Terminal airspace capacity model," *Transportation Research Part A: General*, Vol. 16, No. 4, 1982, pp. 253~260.
- 5) Han, S. H., Yun, T. Y. and Kim, S. H., "Machine Learning Based Capacity Prediction Model of Terminal Maneuvering Area," *Journal of the Korean Society for Aeronautical and Space Sciences*, Vol. 50, No. 3, pp. 215~222.
- 6) Wang, S., Yang, B., Duan, R. and Li, J., "Predicting the Airspace Capacity of Terminal Area under Convective Weather Using Machine Learning," *Aerospace*, October 2023, p. 288
- 7) Nagy, D., "The Lambert Conformal Conic Projection: A Short Algorithm to Calculate the Direct and Inverse Transform," *Critical Incident Stress Management Journal*, Vol. 44, No. 2, 1990, pp. 158~164.
- 8) Kang, J. H. and Lee, H. T., "Development and Implementation of Airspace Coordinate Preprocessing Algorithm," *Proceeding of the Korean Society for Aeronautical and Space Sciences Spring Conference*, April 2022, pp. 155~156.
- 9) Lee, H., Park, B. S. and Lee, H. T., "Waypoint Extraction from Recorded ADS-B Trajectory Data," *The Korean Navigation Institute Conference*, 2016.
- 10) Kenneth, B. S., "An efficient point-in-polygon algorithm," *Computers and Geosciences*, Vol. 4, No. 2, 1978, pp. 173~178.
- 11) Reynolds, D., "Gaussian mixture models," *Encyclopedia of biometrics*, 2015, pp. 827~832.
- 12) Johnson, S. C., "Hierarchical Clustering

Schemes," *psychometrika* 32, 1967, pp. 241~254.

13) Vijaya, Sharma, S. and Batra, N., "Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering," *2019 International Conference on Machine Learning, Big*

Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 568~573.

14) Chandola, V., Banerjee, A. and Kumar, V., "Anomaly Detection: A Survey," *ACM Computing Surveys*, Vol. 41, No. 3, 2009, pp. 1~58.