

가우시안 혼합모델을 이용한 공항 접근 패턴 추출 및 패턴 별 과이탈 확률 분석

Extracting Patterns of Airport Approach Using Gaussian Mixture Models and Analyzing the Overshoot Probabilities

류재영* · 한성민 · 이학태
인하대학교 항공우주공학과

Jaeyoung Ryu* · Seong-Min Han · Hak-Tae Lee

Department of Aerospace Engineering, Inha University, Incheon 22212, Korea

[요 약]

항공기 착륙 시에는 정해진 절차에 따라 접근이 이루어진 다음, 활주로 중심선과 정렬하여 착륙하게 된다. 하지만 공항의 상황, 주변 항공기의 상황, 또는 관제사의 지시 등에 따라 빈번한 레이더 백터링이 일어나기 때문에, 교통 흐름을 파악하거나, 비행 안전성을 파악하기 위해서는 항공기의 접근 패턴을 인지할 필요가 있다. 또한 최종 접근 시 활주로 중심선과 정렬하는 과정에서 과이탈이 발생하는 경우가 있는데, 이는 이후 불안정 접근 등과 같이 보다 위험한 상황을 초래할 수 있다. 본 논문에서는 클러스터링 기법을 이용하여 접근 구간에서의 항공기 궤적들의 패턴을 추출하였다. GMM (Gaussian Mixture Model)을 이용하여 김해공항 접근 항공기 궤적에 대한 클러스터링을 진행하였으며, 2019년 1년간 김해공항으로 착륙한 항공기의 데이터를 이용하였다. 클러스터별 centroid 값을 이용하여, 총 86개의 접근 궤적 패턴을 추출하였다. 그 후 각 클러스터 내 항공기 중 최종 접근시 과이탈하는 항공기를 탐지하여 확률 분포를 계산하였다.

[Abstract]

When an aircraft is landing, it is expected that the aircraft will follow a specified approach procedure and then land at the airport. However, depending on the airport situation, neighbouring aircraft or the instructions of the air traffic controller, there can be a deviation from the specified approach. Detecting aircraft approach patterns is necessary for traffic flow and flight safety, and this paper suggests clustering techniques to identify aircraft patterns in the approach segment. The Gaussian Mixture Model (GMM), one of the machine learning techniques, is used to cluster the trajectories of aircraft, and ADS-B data from aircraft landing at the Gimhae airport in 2019 are used. The aircraft trajectories are clustered on the plane, and a total of 86 approach trajectory patterns are extracted using the centroid value of each cluster. Considering the correlation between the approach procedure pattern and overshoots, the distribution of overshoots is calculated.

Key word : Aircraft trajectories, Air traffic management, Clustering, Gaussian mixture model, Standard terminal area.

<http://dx.doi.org/10.12673/jant.2023.27.6.888>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 5 December 2023; Revised 14 December 2023
Accepted (Publication) 26 December 2023 (29 December 2023)

*Corresponding Author Jaeyoung Ryu

Tel: +82-10-4945-6308

E-mail: jaeyoungRyu@inha.edu

I. 서론

데이터 기반 항공 안전 분석은 항공 산업에서 발생한 다양한 데이터를 수집하고 분석하여 안전문제를 예방하는 것을 목적으로 한다. 최근 항공 전자 장비 발달과 통신 장비의 발달로 인해 데이터 수집이 용이해지면서, 대용량 데이터를 처리하기 위해, 머신러닝을 이용한 연구가 많이 진행되고 있다. 논문[1]과 [2]의 경우는 각각 Extreme Learning machine 알고리즘과 Temporal Logic Learning 알고리즘을 이용하여 항공기 데이터 상에서의 이상 징후를 찾는 연구를 진행하였다. 논문[3]의 경우는 GMM (Gaussian Mixture Model)을 이용하여 항공기 속도, Angle of Attack 등의 파라미터에 대한 이상 징후 탐지 연구를 진행하였다. 머신러닝을 이용하여 항공기 궤적 패턴을 추출하는 연구도 많이 진행되고 있는데, 논문[4]의 경우는 k-means와 DBSCAN 알고리즘을 이용하여 접근 항공기의 패턴을 추출하였다. 논문[5]-[7]의 경우는 GMM 클러스터링을 이용하여 Terminal Airspace 내의 항공기 궤적 패턴을 추출하였으며, 논문[8]의 경우는 DTW와 Agglomerative 알고리즘을 이용하여 접근 항공기의 궤적 패턴을 추출하였다. 논문[9]의 경우는 distance 계산 시, weight를 주고 HDBSCAN을 이용하여 패턴을 추출하였으며, 논문[10]-[11]의 경우는 추출한 패턴을 이용하여 항적을 예측하는 연구를 진행하였다.

본 논문은 GMM 알고리즘을 이용하여 김해공항 접근 항공기들의 접근 궤적 패턴을 추출하고, 패턴 별 과이탈이 일어난 확률을 계산한다. 2019년 김해 공항으로 착륙한 항공기의 데이터를 이용하였으며, GMM 알고리즘의 k값을 구하기 위해 Agglomerative 알고리즘을 이용하였다. 총 53572대의 항공기 데이터를 이용하여, 86개의 클러스터로 분류하였으며 각각 클러스터의 중심값을 이용하여, 접근 궤적 패턴을 추출하였다. 클러스터의 중심값 궤적을 확인한 결과, 유사한 중심값 궤적을 가지는 클러스터들이 존재하였기 때문에, 한 단계 큰 클러스터로 통합하였으며, 각 클러스터 별, 과이탈이 일어난 확률을 계산하였다. 계산 결과 MASTA를 접근 단계의 시작 지점으로 하여, 활주로 36으로 착륙하는 항공기와 KALEK을 시작 지점으로 하여 활주로 36으로 착륙하는 항공기들의 경우, 과이탈이 일어난 확률이 높았으며, 특히 활주로 정렬 시, 정렬 거리가 짧을수록 과이탈이 일어난 확률이 높은 것을 확인하였다.

본 논문의 구성은 다음과 같다. II장에서는 데이터 전처리 방법을 설명한다. III장에서는 접근 단계 항적 패턴을 추출하기 위하여 사용한 클러스터 알고리즘 및 데이터 Matrix를 설명하며, IV장에서는 클러스터링 결과로 나온 중심값 궤적 패턴을 제시하고, 클러스터 별, 과이탈이 일어난 확률에 대해 설명한다. V장에서는 본 논문의 결론과 향후 연구 계획을 제시한다.

II. ADS-B 기반 항적 데이터 재생성

II장에서는 ADS-B 데이터에 대한 설명과 머신러닝에 적용하기 위한 전처리 방법 및 과이탈 계산 방법에 대해 설명한다.

2.1 ADS-B 데이터

ADS-B 데이터는 transponder를 장착한 항공기가 자신의 비행 정보를 송출할 경우, 지상국 또는 근처의 항공기가 이를 수신하며 서로의 위치 정보를 공유하는 데이터이다. 항공기의 위치, 속도, 고도, 방향 등의 정보를 가지고 있으며, 2020년부터 미국과 유럽의 모든 민간 항공기는 ADS-B 시스템 장착을 의무화하였다 [12]. 그림 1은 ADS-B 데이터 예시를 나타내며, 본 논문에서는 FlightAware에서 구매한 2019년도 ADS-B 데이터를 이용하여 연구를 진행하였다.

*Id	Callsign	Type	Ogn	Dst	TimeSnap	LAT	LON	ALT	GSD	CRS
509458560	KAL1953	B738	RKTU	RKPC	1540695671	36.69629	127.4267	4025	148	281
509458560	KAL1953	B738	RKTU	RKPC	1540695677	36.69694	127.4225	4075	158	280
509458560	KAL1953	B738	RKTU	RKPC	1540695687	36.69836	127.4128	4250	174	280
509458560	KAL1953	B738	RKTU	RKPC	1540695697	36.69983	127.4021	4425	190	279

그림 1. ADS-B 데이터 포맷
Fig. 1. ADS-B Trajectory Data Format

2.1.2 클러스터링을 위한 데이터 전처리

접근 단계의 항공기의 궤적을 분석하기 위해서는 우선 접근 단계를 설정할 필요가 있다. 일반적으로 항공기는 계기비행 시에 항로를 비행하여 목적지 부근에 이르러 항로에서 공항으로 전환하기 위한 경로를 이용한다. 대표적으로 Standard Terminal Arrival Route (STAR)를 이용하며, 여러 항로에서부터 전환경로를 통하여 계기접근을 유도한다. 접근 단계의 시작 지점을 STAR 절차 첫 지점으로부터 10 nmi 내에 들어온 경우로 지정하였으며, 접근 단계의 마지막 지점은 공항으로부터 3 nmi 내에 들어올 경우로 지정하였다. 2019년도 ADS-B 데이터를 사용하였기 때문에, 2019년도 김해공항의 STAR 절차를 이용하여 접근 단계를 지정하였으며, 그림 2는 2019년 김해공항 STAR 시작 지점과 김해공항을 나타낸다. STAR의 시작 지점을 빨간 원으로 표현하였으며, 김해공항을 노란원으로 표현하였다. 표 1의 경우는 STAR 시작 지점에 파라미터를 나타내었다.

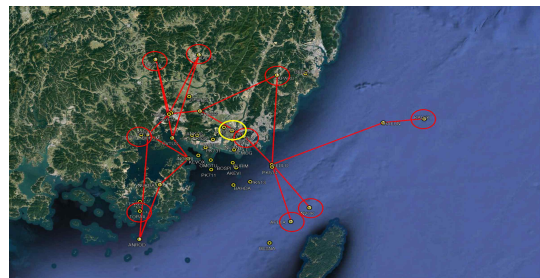


그림 2. 김해공항 STAR 시작지점
Fig. 2. RKPK first STAR start WPT

표 1. 김해공항 STAR 시작지점 파라미터

Table. 1. RPKK STAR first WPT parameter

STAR wpt	Latitude	Longitude	Distance to RPKK (nmi)
APARU	35.4117	129.1589	17.6420
KALEK	35.2088	129.8847	46.5855
INVOK	34.7886	129.3231	30.1397
APELA	34.7231	129.2333	30.9760
PSN	35.1226	128.9996	4.5545
MASTA	35.4797	128.5611	25.8077
KALDO	35.5034	128.7740	21.0089
SARAM	35.1268	128.5296	20.3404
TOPAX	34.7654	128.4977	32.9658

ADS-B 데이터와 STAR 정보를 담고 있는 AIP 문서의 경우, 위치 정보를 위도, 경도로 나타내기 때문에 이를 평면상에 투영할 필요가 있다. 본 연구에서는 김해공항의 좌표를 원점으로 하는 람베르트 정각원추도법을 사용하여 평면상에 투영하였으며 사용한 지표는 표 2에 나타내었다 [13].

표 2. 람베르트 정각원추도법 기준 좌표

Table. 2. Lambert conic conformal projection values

Threshold	Standard Parallel 1	Standard Parallel 2	Origin Position
latitude	35.1794°	38.17944°	35.1794°
longitude	126.9380	130.9380°	128.9380°

ADS-B 데이터의 경우 1초 또는 그 이하의 주기로 항공기 데이터를 방송하는 것을 목적으로 한다. 하지만 항공기 내 송신기의 상태, 또는 지상국 수신기의 상태 등에 다양한 이유로 데이터가 기록되지 않는 경우가 존재하며, 본 연구에서 사용한 ADS-B 데이터 또한 일정 주기를 가지지 않는다. 이를 보완하기 위해 항적 데이터를 1초 단위로 시간 동기화를 진행하였다. 위에 제시한 람베르트 정각원추도법을 이용하여 항공기 위도, 경도 좌표를 (x, y) 값으로 나타냈고 선형보간을 통해 1초 간격으로 데이터를 동기화 하였다. 그림 3은 ADS-B 데이터를 이용하여 재생성한, 김해공항으로 접근하는 항공기 데이터의 예시를 보여준다. 항공기 항적은 검정색 점선으로 표현하였으며, 김해공항 STAR 절차를 가시화하였다. STAR 시작 지점을 통과한 다음, 착륙할 때까지의 시간이 비정상적인 데이터를 제외한 총 53572대의 항공기 데이터를 재생성하였다.

2.2 과이탈 판별 알고리즘

과이탈은 항공기가 활주로와 정렬하는 과정 중 활주로를 교차한 다음, 일정 범위 이상으로 이탈하는 현상을 의미한다. 이는 착륙 시에 날씨 조건, 항공기의 제동 장치 불량 또는 작동 이상 등의 다양한 이유로 나타날 수 있으며, 활주로 이탈 또는 경차착륙을 유발할 수 있기 때문에, 사전에 예방할 필요가 있다.

접근 절차 패턴과 과이탈 사이의 상관관계가 존재할 것으로 예상하여, 접근 절차 패턴별 과이탈이 일어난 확률을 분석하였다.

본 연구에서는 논문[14]에서 제시한 최종접근경로 과이탈 산출 알고리즘을 이용하여 과이탈 계산을 진행하였다. 과이탈 산출 알고리즘에는 활주로 판별 알고리즘과 과이탈 발생 시, 최대 과이탈 거리를 산출하는 알고리즘이 포함되어 있다. 그림 4는 과이탈 거리 계산 방법을 가시화한 그림이다. 먼저 착륙을 준비하는 항공기의 3D 궤적을 평면상으로 투영하고, 각 데이터의 위치에서 활주로 중심선을 연장한 선 사이의 거리를 계산한다. 해당 거리의 절대값이 100 m ~ 1000 m일 경우 과이탈이 일어난 것으로 판단하였다. 과이탈 거리는 식 (1), (2)과 같이 계산한다. \vec{B}_{air} 는 도착 runway threshold와 항적 사이의 벡터이고, \vec{A}_{run} 는 runway threshold의 방향 벡터를 의미한다.

$$\sin(\theta) = \frac{|\vec{B}_{air} \times \vec{A}_{run}|}{|\vec{A}_{run}| |\vec{B}_{air}|} \tag{1}$$

$$d = |\vec{B}_{air}| \sin(\theta) \tag{2}$$

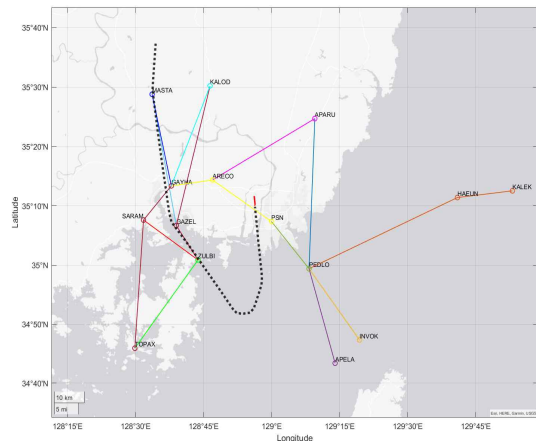


그림 3. 궤적 예시 : KAL 1105
Fig. 3. Trajectory sample : KAL 1105

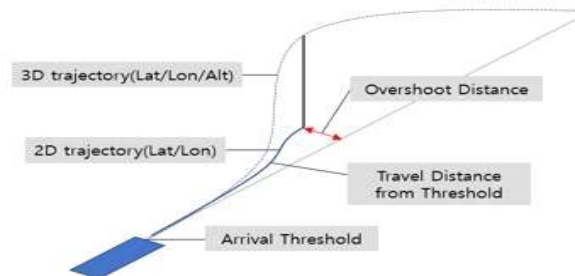


그림 4. 과이탈 계산 과정
Fig. 4. Overshoot calculation process

III. 클러스터링 알고리즘

III장에서는 본 연구에서 사용한 클러스터링 알고리즘에 대해 설명한다.

3-1 Data Matrix

클러스터링을 진행하기 위해서는 Data Matrix를 구성해야 한다. Data Matrix는 n 개의 데이터가 p 개의 차원을 가지는 경우 ($n \times p$) 행렬을 가지게 된다. n 은 항공기 궤적 수를 의미하고, p 는 항공기 궤적의 feature dimension을 의미한다. 본 연구에서는 평면상에서의 항공기 궤적을 이용하여 클러스터링을 진행하였다.

$$X_n = [x_1^1, x_1^2, \dots, x_1^f, \dots, x_t^1, x_t^2, \dots, x_t^f] \quad (3)$$

사용한 데이터는 식 (3)처럼 표현할 수 있으며 X_n 은 n 번째 항공기 데이터를 의미하고, t 는 데이터의 시간을 의미한다. f 는 데이터의 feature를 의미하며, 항공기의 위도, 경도, 고도, 속도 등에 파라미터들이 들어갈 수 있다. 평면상에서 클러스터링을 진행하였기 때문에, 항공기의 위도, 경도를 람베르트 정각원 추도법으로 투영하여 얻은 km 단위의 (x, y) 좌표를 이용하여 feature dimension을 구성하였다. Timestep은 2500으로 구성하였으며, 1초 간격으로 시간 동기화를 진행하였기 때문에, 접근 단계 시작 지점을 통과한 기준으로 2500초 동안의 항공기 궤적 데이터를 이용하여 Data Matrix를 구성하였다. 만약 항공기가 2500초 이전에 접근 단계 마지막 지점을 통과한 경우, 검색 공항의 (x, y) 좌표인 원점을 사용하여 Data Matrix를 구성하였다.

3-1 Agglomerative Clustering 알고리즘[15]

Agglomerative Clustering 알고리즘은 각 데이터가 모두 나뉜 상태에서, 작은 단위로부터 클러스터링을 시작하여 모든 데이터를 하나의 클러스터로 묶을 때까지 반복하는 알고리즘이다. 원하는 클러스터 수 또는 일정 문턱 거리에서 클러스터링을 끊을 수 있기 때문에, 클러스터의 수를 알지 못할 때 사용하기에 효과적이다. 그림 5는 Agglomerative Clustering을 진행하여 얻을 수 있는 Dendrogram을 나타내었다. 본 연구에서는 GMM을 이용하여 항공기 궤적 분석을 진행하는데, 이를 위한 k 값을 얻기 위해 해당 알고리즘을 사용한다.

Agglomerative 알고리즘을 진행하기 위해서는 각 클러스터 별 문턱 거리를 구해야 한다. 문턱 거리는 노이즈나 이상치에 덜 민감한 장점이 있는 ward 연결법을 이용하였다[16]. 식 (4)는 ward 연결법 구하는 식을 의미한다. Timestep을 2500으로 잡았고, 평면상에 좌표인 (x, y) 를 feature로 삼았기 때문에 p 는 5000의 값을 가지게 된다. 각 feature dimension 별 문턱 거리를

1로 설정하였으며 총 문턱거리를 5000으로 하여 클러스터링을 진행하였다.

$$d(A, B) = \frac{\|\mu_A - \mu_B\|^2}{\frac{1}{n_A} + \frac{1}{n_B}} \quad (4)$$

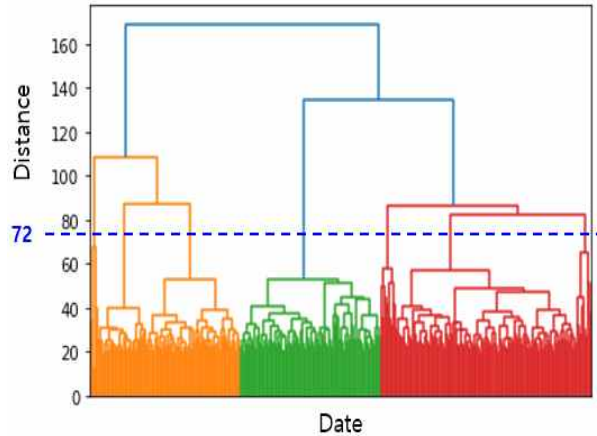


그림 5. 덴드로그램 : 예시
Fig. 5. Dendrogram : Example

3.3 GMM Clustering 알고리즘[17]

GMM 알고리즘은 분석 대상 데이터들이 가우시안 분포를 가진다는 가정에서 시작한다. 하나의 가우시안 분포를 따르기에는 데이터의 경향성이 일정하지 않기 때문에 여러 개의 혼합된 분포를 따른다고 가정하여, 각각의 가우시안 분포를 찾는 것을 목적으로 한다. GMM 알고리즘은 클러스터의 평균값을 의미하는 μ_k , 분산값을 의미하는 Σ_k , 전체 분석 대상 데이터에서 각각의 가우시안 분포의 영향력을 의미하는 π_k 로 구성된다. 각 데이터의 확률 밀도 함수는 식 (5)와 같이 계산하며, 전체 데이터로부터 특정 통계 모델의 적합성을 확인하는 Likelihood 함수는 식 (6)과 같이 계산한다. GMM은 Likelihood 함수를 최대화하는 μ_k, Σ_k, π_k 를 찾는 것을 목적으로 한다.

$$p(x) = \sum_{k=1}^K \pi_k N(x | \mu_k, \Sigma_k) \quad (5)$$

$$P(X | \theta) = \prod_{i=1}^N \left(\sum_{k=1}^K \pi_k N(x_i | \mu_k, \Sigma_k) \right) \quad (6)$$

식 (6)을 바로 사용하는 경우, 미분 계산이 힘들기 때문에 식 (6)에 로그를 붙인 Log-likelihood를 사용하며, 이를 최대화하기 위해 Expectation-Maximization (EM) 알고리즘을 사용한다. EM 알고리즘은 데이터 x_n 이 특정 분포에 속할 확률을 계산하는 단계인 E단계와, Log-likelihood를 최대화하는 모수를 추정하는 M단계로 나뉜다. E단계에서는 데이터가 특정 분포에 속할

확률인 Responsibility를 식 (7)과 같이 계산한다.

$$\gamma(z_k) = p(z_k = 1 | x_i) = \frac{\pi_k N(x_i | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j N(x_i | \mu_j, \Sigma_j)} \quad (7)$$

M단계는 Log-likelihood를 최대화하는 μ_k, Σ_k, π_k 를 찾는 과정이다. Log-likelihood 함수는 위로 볼록한 모양을 갖기 때문에 편미분을 진행하여 미분값이 0인 방정식을 푸는 정규방정식을 사용하여 최대화를 진행한다. 식(8)-(10)처럼 평균, 분산, 확률값을 구할 수 있다. μ_k, Σ_k, π_k 가 수렴할 때까지 EM 알고리즘을 반복한다.

$$\mu_k = \frac{\sum_{n=1}^N \gamma(z_{nk}) x_n}{\sum_{n=1}^N \gamma(z_{nk})} \quad (8)$$

$$\Sigma_k = \frac{\sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^N \gamma(z_{nk})} \quad (9)$$

$$\pi_k = \frac{1}{N} \sum_{n=1}^N \gamma(z_{nk}) \quad (10)$$

IV. 클러스터링 결과 및 과이탈 분포

IV장에서는 GMM을 이용한 항공기 궤적 클러스터링 결과, 클러스터 별 과이탈 분포를 제시한다.

4-1 클러스터링 결과

Agglomerative 알고리즘을 이용하여 클러스터링을 진행한 결과 86개의 클러스터로 구분되었기 때문에 GMM 알고리즘을 실행할 때, 클러스터 수를 의미하는 k값을 86으로 설정하였다. 그림 6은 GMM 알고리즘 계산 결과, 86개의 클러스터 중 1% 이상의 항적을 가지는 클러스터의 중심값을 나타내며, 표 3은 GMM 알고리즘을 통한 클러스터 별 항공기 대수를 나타낸다. 클러스터 중 가장 많은 항적 데이터를 가지는 43번과 75번 항적을 그림 7과 8에 나타내었다. 노란실선은 각각의 항공기 궤적을 의미하며, 빨간선은 클러스터의 중심값, 파란선은 클러스터의 분산값을 가시화하였다.

클러스터 43의 경우는 TOPAX를 접근절차의 시작으로 하는 항공기의 클러스터이지만, APELA, INVOK를 시작지점으로 하는 항공기도 존재하였다. APELA, INVOK를 접근절차의 시작으로 하는 다른 클러스터의 경우, 활주로와 항공기가 정렬하기 전, 항공기 선회 기동이 존재하였다. 하지만 클러스터 43 내

에 존재하는 해당 항공기들은 활주로와 정렬 전까지, 직선으로 기동하는 특징이 있으며, 다른 항공기들보다 빠르게 착륙하는 것을 확인하였다. 항공기가 착륙한 이후, 김해공항의 좌표인 원점으로 Data matrix를 구현하였기 때문에, 착륙 시간이 비교적 짧은 TOPAX를 시작 지점으로 하는 클러스터에 포함되었으며, 총 7268 궤적 중, 427대의 항공기가 해당 경향을 보였다.

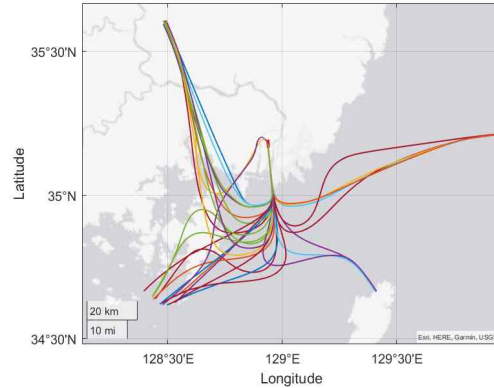


그림 6. 김해공항 : 클러스터 중심값
Fig 6. RPKK : Cluster Centroid

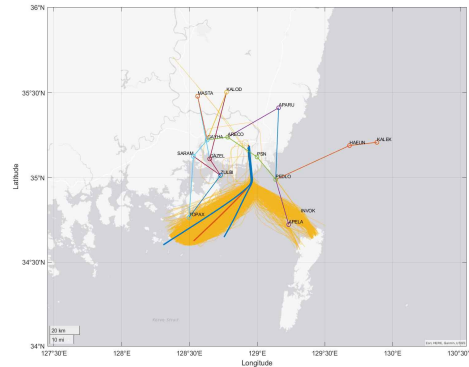


그림 7. 김해공항 : 클러스터 43번 궤적
Fig. 7. RPKK : Cluster 43 trajectories

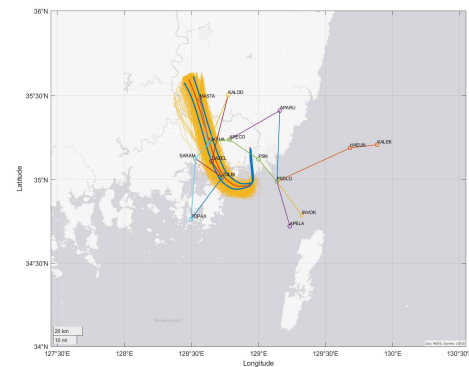


그림 8. 김해공항 : 클러스터 75번 궤적

Fig. 8. RPKK : Cluster 75 trajectories

표 3. 클러스터 별 항공기 수

Table. 3. Number of Aircraft in each Cluster

Cluster	Number	Cluster	Number	Cluster	Number	Cluster	Number
cluster 0	357	cluster 22	72	cluster 44	190	cluster 66	1
cluster 1	1523	cluster 23	150	cluster 45	95	cluster 67	560
cluster 2	612	cluster 24	3324	cluster 46	452	cluster 68	1070
cluster 3	1199	cluster 25	152	cluster 47	74	cluster 69	1569
cluster 4	303	cluster 26	1443	cluster 48	230	cluster 70	1534
cluster 5	872	cluster 27	319	cluster 49	391	cluster 71	710
cluster 6	742	cluster 28	602	cluster 50	430	cluster 72	299
cluster 7	547	cluster 29	273	cluster 51	2227	cluster 73	47
cluster 8	1472	cluster 30	665	cluster 52	96	cluster 74	482
cluster 9	1310	cluster 31	86	cluster 53	299	cluster 75	299
cluster 10	256	cluster 32	233	cluster 54	107	cluster 76	705
cluster 11	305	cluster 33	337	cluster 55	786	cluster 77	444
cluster 12	318	cluster 34	122	cluster 56	155	cluster 78	91
cluster 13	67	cluster 35	689	cluster 57	1065	cluster 79	336
cluster 14	209	cluster 36	1555	cluster 58	216	cluster 80	224
cluster 15	738	cluster 37	1268	cluster 59	90	cluster 81	68
cluster 16	1378	cluster 38	356	cluster 60	194	cluster 82	708
cluster 17	1147	cluster 39	57	cluster 61	47	cluster 83	633
cluster 18	279	cluster 40	81	cluster 62	61	cluster 84	1758
cluster 19	425	cluster 41	163	cluster 63	228	cluster 85	105
cluster 20	132	cluster 42	80	cluster 64	445	Total	53572
cluster 21	195	cluster 43	7268	cluster 65	380		

한편 STAR 시작지점, 착륙하는 활주로 방향 등에 따라 유사한 형태의 중심값을 가지는 클러스터들이 존재하였으며 이들의 경우, 착륙하는 시간에 따라 분류 되는 것을 확인하였다. 본 연구에서는 시간대 별 항공기의 위치 좌표로 feature dimension을 구분하였기 때문에, 항공기 속력값을 feature 값으로 사용하지 않아도 클러스터 내에서 자동적으로 분류되었다.

유사한 중심값 궤적을 가지는 클러스터들을 다시 하나의 클러스터로 통합하였으며, 86개의 클러스터는 전체적으로 15가지 궤적 패턴과 noise 패턴으로 구분되었다. 전체적으로 구분된 15가지 패턴을 빅클러스터, 그 안에 존재하는 클러스터를 스몰

클러스터로 부르도록 한다.

빅클러스터는 항공기의 접근 절차가 시작되는 지점으로 가장 먼저 나누었졌으며, 활주로 방향, 선회비행 여부 등에 따라서 나누었다. TOPAX, KALEK, APELA, INVOK를 시작 지점으로 하는 접근 단계의 경우, 선회비행을 하는 항적이 많이 존재하였기 때문에 이를 따로 구분하였다. 그림 9와 10은 TOPAX를 접근 단계의 시작 지점으로 하고 활주로 36으로 착륙하는 항적들의 중심값과 중심값의 시간대 별 위,경도를 나타낸다. 그림 11과 12는 TOPAX를 시작으로 하여 중간에 ZULBI를 통과하고, 활주로 36으로 착륙하는 항적에 대해 나타낸다.

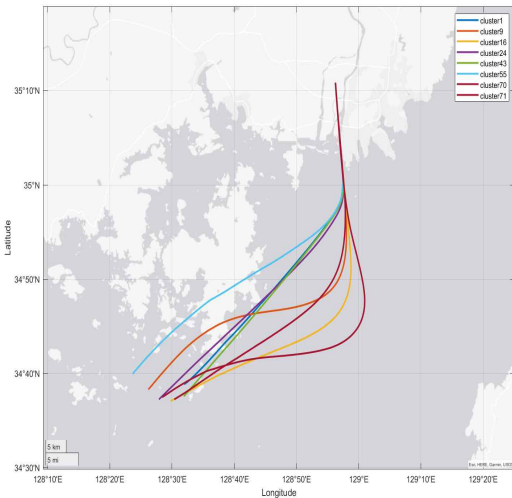


그림. 9. TOPAX를 지나 런웨이 36으로 들어가는 궤적의 중심값
Fig. 9. Centroid of Trajectories through TOPAX and into Runway 36

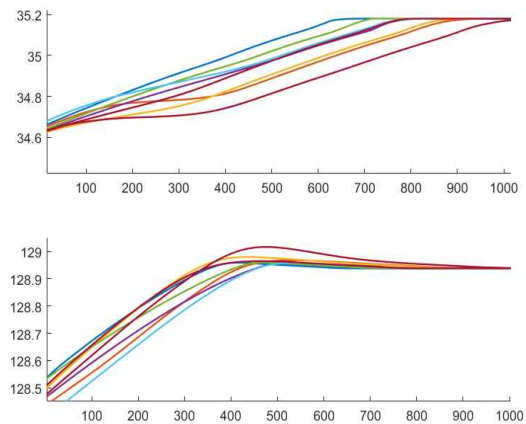


그림. 10. TOPAX를 지나 런웨이 36으로 들어가는 궤적 중심값의 위도와 경도
Fig. 10. Latitude and Longitude of the Trajectory Centroid through TOPAX and into Runway 36

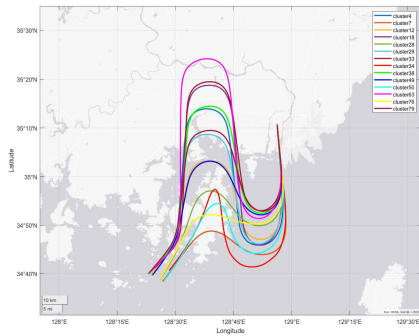


그림 11. TOPAX와 ZULBI를 통해 런웨이 36으로 들어가는 궤적의 중심값

Fig. 11. Centroid of Trajectories through TOPAX and ZULBI and into Runway 36

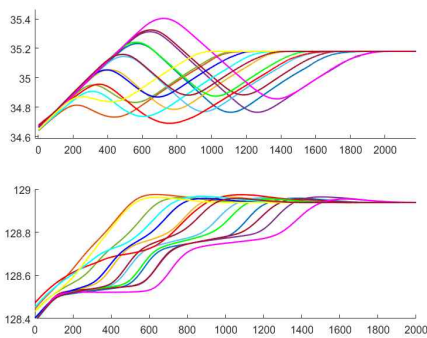


그림 12. TOPAX, ZULBI를 통해 런웨이 36으로 들어가는 궤적 중심값의 위도와 경도

Fig. 12. Latitude and Longitude of the Trajectory Centroid through TOPAX and ZULBI and into Runway 36

그림 9에 나타낸 TOPAX를 접근 단계 시작 지점으로 하고 활주로 36으로 착륙하는 클러스터의 중심값을 살펴보면, 클러스터 1, 24, 43의 경우, 유사한 궤적을 가지는 것을 볼 수 있다. 하지만 그림 10에 나타낸 시간에 따른 중심값의 위도와 경도를 비교한 경우, 클러스터 1이 가장 빠르게 착륙하였으며, 그 이후 43번, 24번 순으로 착륙한 것을 볼 수 있다. 그림 11을 보면 항적 별 ZULBI를 통과하기 위한 선회 위치에 따라 클러스터가 구분되는 것을 볼 수 있으며, 클러스터 12번과 49번, 18번과 33, 29번과 79번이 유사한 형태를 가지고 있지만 이들 또한 착륙하는 시간에 따라 구분되는 것을 볼 수 있다.

4.2 클러스터별 과이탈 확률

전체 항적 대비 6.85퍼센트의 항적에서 과이탈이 일어났으며, 빅클러스터와 스몰 클러스터 별 과이탈이 일어난 확률을 분석하였다. 표 4은 빅클러스터 별 과이탈이 일어난 확률을 나타낸다.

빅클러스터들 중 TOPAX를 시작 지점으로 하여 곧바로 활주로 36으로 들어가는 항공기의 경우, 6.8퍼센트의 확률로 과

이탈이 일어났으며, ZULBI를 통과한 다음 활주로 36으로 들어가는 경우는 0.8퍼센트의 확률로 과이탈이 일어났다. 활주로 18을 이용하는 경우는 6.7퍼센트의 확률로 과이탈이 일어났으며, TOPAX에서 접근 단계를 시작할 경우, 착륙한 활주로보다 ZULBI 통과 여부에 따라 확률이 달라지는 것을 볼 수 있다.

MASTA를 시작 지점으로 하여, 활주로 36으로 착륙하는 항공기의 경우 9.4퍼센트의 확률로 과이탈이 일어났으며 KALEK을 시작 지점으로 하여 활주로 36으로 착륙하는 항공기들의 경우는 10.8퍼센트의 확률로 과이탈이 일어났다. 그림 13은 MASTA에서 활주로 36으로 착륙하는 스몰 클러스터들의 중심값, 그림 14는 KALEK에서 활주로 36으로 착륙하는 스몰 클러스터들의 중심값을 나타낸다.

표 4. 빅클러스터 별 과이탈 발생 비율

Table 4. Overshoot Rate per Big Cluster

Big Cluster	Overshoot (%)	Big Cluster	Overshoot (%)
TOPAX , Runway 36	6.80	TOPAX, Turn around , Runway 36	1.14
TOPAX , Runway 18	6.73	TOPAX , ZULBI, Runway 36	0.84
MASTA , Runway 36	9.40	MASTA, Runway 18	3.28
APELA_INVOK, Runway 36	2.89	APELA_INVOK, Turn around, Runway 36	2.37
APELA_INVOK, Runway 18	5.89	KALEK, Runway 36	10.80
KALEK, Turn around, Runway 36	2.72	KALEK, Runway 18	5.56
APARU, Runway 36	3.97	APARU, Runway 18	0
SARAM, Runway 36	6.47	Total	6.85

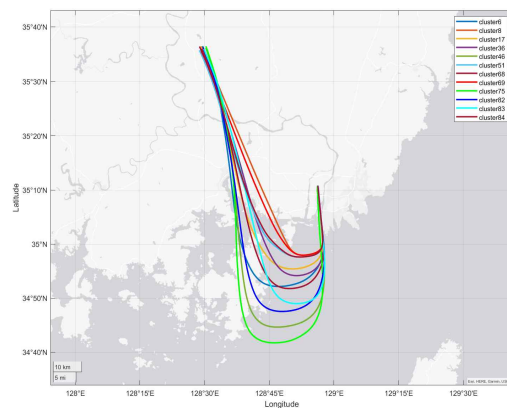


그림 13. MASTA를 지나 런웨이 36으로 들어가는 궤적의 중심값

Fig. 13. Centroid of Trajectories through MASTA and into Runway 36

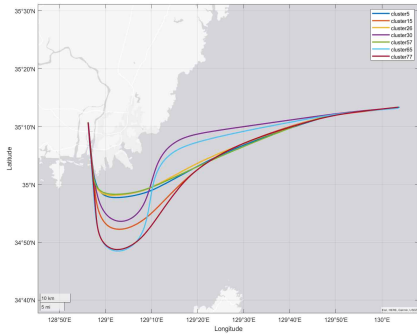


그림 14. KALEK를 지나 런웨이 36으로 들어가는 궤적의 중심값
 Fig. 14. Centroid of Trajectories through KALEK and into Runway 36

MASTA를 접근 단계의 시작 지점으로 하고, 활주로 36으로 착륙하는 스몰 클러스터 중에, 주황색, 하늘색, 빨간색으로 표현된 클러스터 8, 51, 69의 경우는 궤이탈이 일어난 확률이 13 퍼센트보다 높았다. 이들은 다른 중심값의 궤적보다 활주로의 정렬하기 전, 선회한 길이가 짧았으며, 이는 선회한 길이가 짧을수록 궤이탈이 일어난 확률이 높은 것을 보여준다. KALEK을 접근 단계의 시작 지점으로 하고, 활주로 36으로 착륙하는 경우도 비슷한 경향을 가졌다. 파란색, 연두색, 노랑색으로 표현된 클러스터 5, 26, 57번의 경우 궤이탈 비율이 10퍼센트보다 높았는데, 이들 또한 활주로의 정렬 전, 선회하는 길이가 다른 클러스터 중심값 궤적에 비해 짧은 것을 볼 수 있다. 이 중 클러스터 5는 10퍼센트, 클러스터 26, 57의 경우 15퍼센트, 20퍼센트의 확률로 궤이탈이 일어났는데, 클러스터 5의 경우, 26, 57보다 약간 더 멀리서 선회하는 것을 볼 수 있다. 이 둘의 결과로 보아, 활주로의 정렬하기 전, 선회하는 길이와 궤이탈 사이의 연관관계가 있다고 볼 수 있다.

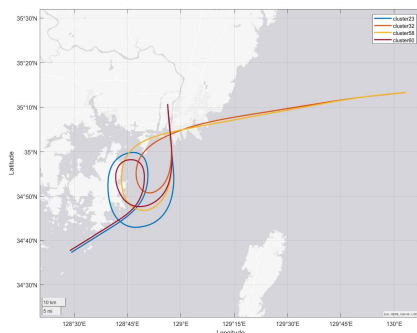


그림 15. 선회비행 후, 런웨이 36으로 들어가는 궤적의 중심값
 Fig. 15. Centroid of Trajectories Where the Circling Flight Occurred

선회비행을 진행한 빅클러스터의 경우, 궤이탈이 일어나는 비율이 다른 빅클러스터에 비해 적었는데, 이는 선회비행을 하면서 활주로의 정렬을 미리 준비 하기 때문에 일어나는 현상

으로 보인다. 그림 15는 TOPAX와 KALEK을 시작지점으로 하고, 선회비행을 진행한 후, 활주로 36으로 들어오는 클러스터의 중심값을 보여준다.

V. 결론

본 연구는 김해공항 착륙 항공기를 대상으로, 접근 단계 시의 항공기 궤적 패턴을 확인하였으며, 접근 단계 패턴 별, 궤이탈이 일어난 확률을 분석하였다. 머신러닝 기반의 데이터 군집화 모델인 GMM 알고리즘을 이용하였으며, GMM 알고리즘 계산 시, 필요한 k값은 Agglomerative 알고리즘을 이용하여 구하였다. 그 결과 86개의 클러스터를 구하였으며, 클러스터의 중심값을 이용하여 접근 궤적 패턴을 추출하였다. 클러스터의 중심값들은 유사한 궤적을 가지는 경우가 존재하였기 때문에, 접근 단계 시작 지점, 활주로 방향 등에 따른 빅클러스터로 구분하였다. 클러스터 별 궤이탈 발생 여부를 계산하였으며, 궤이탈이 일어난 확률을 분석하였다. 분석 결과, MASTA, KALEK을 접근 단계의 시작 지점으로 하고, 활주로 36으로 착륙하는 경우, 궤이탈이 많이 일어났으며, 특히 선회 길이가 짧을수록 궤이탈이 많이 일어나는 것을 확인하였다.

본 논문을 통해 수행한 연구 결과를 바탕으로, 국내 주요 공항들에 접근 단계 패턴화를 진행할 것이며, 궤이탈이 아닌 다른 안전문제와도 연관지어 연구를 진행할 계획이다.

Acknowledgments

본 논문은 2023년도 정부(국토교통부)의 재원으로 ‘항공안전 빅데이터 분석기술 및 빅데이터 기반 항공안전 분석플랫폼 개발 (과제번호 : RS-2020-KA158275)’ 연구의 지원을 받아 수행한 연구로서, 관계 부처에 감사드립니다.

References

- [1] V. M. Janakiraman and D. Nielsen, “Anomaly detection in aviation data using extreme learning machines,” *2016 International Joint Conference on Neural Networks (IJCNN)*, Vancouver: BC, pp. 1993-2000, 2016.
- [2] R. Deshmukh, D. Sun, K. Kim, and I. Hwang, “Temporal logic learning-based anomaly detection in metroplex terminal airspace operations,” *Transportation Research Part C: Emerging Technologies*, Vol. 126, May. 2021.
- [3] L. Li, R.J. Hansman, R. Palacios, R. Welsch, “Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring,” *Transportation Research Part C Emerging*

Technologies, Vol. 64, pp. 45-57, Mar. 2016.

- [4] G. Xu hao, Z. Jun feng, and P. Zihan, "Trajectory clustering for arrival aircraft via new trajectory representation," *Journal of Systems Engineering and Electronics*, Vol. 32, pp. 473-486, Apr. 2021.
- [5] M. C. R. Murça and M. D. Oliveira, "A Data-Driven Probabilistic Trajectory Model for Predicting and Simulating Terminal Airspace Operations," *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, San Antonio: TX, pp. 1-7, 2020.
- [6] S. T. Barratt, M. J. Kochenderfer, and S. P. Boyd, "Learning Probabilistic Trajectory Models of Aircraft in Terminal Airspace From Position Data," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 9, pp. 3536-3545, Sep. 2019.
- [7] W. Zeng, Z. Xu, Z. Cai, X. Chu, and X. Lu, "Aircraft Trajectory Clustering in Terminal Airspace Based on Deep Autoencoder and Gaussian Mixture Model," *Aerospace*, pp. 266, Aug. 2021.
- [8] C. Deng, K. Kim, H. -C. Choi, and I. Hwang, "Trajectory Pattern Identification for Arrivals in Vectored Airspace," *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*, San Antonio, TX, USA, pp. 1-8, 2021.
- [9] S. J. Corrado, T. G. Puranik, O. J. Pinon, and D. N. Mavris, "Trajectory Clustering within the Terminal Airspace Utilizing a Weighted Distance Function," *Proceedings*, Vol. 59, No. 7, 2020.
- [10] C. Deng, H. -C. Choi, H. Park, and I. Hwang, "Trajectory pattern identification and classification for real-time air traffic applications in Area Navigation terminal airspace," *Transportation Research Part C: Emerging Technologies*, Vol. 142, pp. 103765, 2022.
- [11] H. -C. Choi, C. Deng, and I. Hwang, "Hybrid Machine Learning and Estimation-Based Flight Trajectory Prediction in Terminal Airspace," in *IEEE Access*, Vol. 9, pp. 151186-151197, 2021.
- [12] J. Sun, The 1090 Megahertz Riddle [Internet]. Available: <https://mode-s.org/decode/content/ads-b/1-basics.html>
- [13] D. Nagy, "The Lambert Conformal Conic Projection: A Short Algorithm to Calculate the Direct and Inverse Transform" *CISM Journal*, Vol. 44, No. 2, pp. 158-164, 1990.
- [14] S. M. Han, B. S. Park, and H. T. Lee, "Development of Final Approach Overshoot Calculation Algorithm," *Proceeding of the 2023 KSAS Spring Conference*, pp. 1138-1139, 2023.
- [15] S. C. Johnson, "Hierarchical Clustering Schemes," *Psychometrika*, Vol. 32, No. 3, pp. 241-254, 1967.
- [16] S. Sharma., N. Batra, "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)*, pp. 568-573, 2019.
- [17] D. A. Reynolds, "Gaussian mixture models.," *Encyclopedia of biometrics*, Vol. 741, pp. 827-832, 2009.



류 재 영 (Jae-young Ryu)

2020년 8월 : 인하대학교 항공우주공학과 (공학사)
2020년 9월 ~ 2022년 8월 : 인하대학교 항공우주공학과 석사과정
2022년 9월 ~ 현재: 인하대학교 항공우주공학과 박사과정
※ 관심분야 : 항공교통, M&S



한 성 민 (Seong-Min Han)

2022년 2월 : 인하대학교 항공우주공학과 (공학사)
2022년 3월 ~ 현재: 인하대학교 항공우주공학과 석사과정
※ 관심분야 : 항공교통, 데이터 분석



이 학 태 (Hak-Tae Lee)

2006년 1월 : 미국 스탠포드대학교 항공우주공학과 (공학박사)
2013년 9월 ~ 2017년 8월 : 인하대학교 항공우주공학과 조교수
2017년 9월 ~ 2022년 9월 : 인하대학교 항공우주공학과 부교수
2022년 9월 ~ 현재 : 인하대학교 항공우주공학과 정교수
※ 관심분야 : 항공교통, 항공기 공력 설계