creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

도시계획학석사 학위논문

공군 비행단 지원을 위한 AI 기반 급변풍 예측 모델 개발

2022년 1월

서울대학교 환경대학원

환경계획학과 환경관리학전공

곽 준 영

공군 비행단 지원을 위한 AI 기반 급변풍 예측 모델 개발

지도교수 정 수 종

이 논문을 도시계획학석사 학위논문으로 제출함
2022년 1월

서울대학교 환경대학원
환경계획학과 환경관리전공
곽 준 영

곽준영의 석사 학위논문을 인준함
2022년 1월

위 원 장 오 능 환 (인)

부위원장 손 석 우 (인)

위 원 정 수 종 (인)

초 록

항공분야에서 급변풍은 안전에 위협을 주는 가장 큰 이슈 중 하나이다. 급변풍은 급격한 바람의 변화로, 항공기 안전에 보이지 않는 큰 위협을 준다. 현재 급변풍에 대한 정보는 민간공항에만 제공되고 있고, 공군 비행단에는 급변풍 탐지 장비조차 설치되어 있지 않은 실정이다. 군용기는 임무 특성에 따라 기상 악조건 하에도 이착륙을 강행해야 하기 때문에, 군 공항에 대한 급변풍 정보의 확보가 시급하다. 이에 본 연구는 민간공항에서 제공되는 급변풍 정보를 토대로 전국에 두루 적용할 수 있는 급변풍 예측 AI 모델(Artificial Intelligence, 인공지능)을 개발하여, 급변풍 관측 장비가 없는 군 공항에서도 급변풍 정보를 확보하고자 하였다. 과거 데이터를 축적하고, 관측 장비가 설치돼 있는 전국 3개의 민간공항(인천, 제주, 양양)을 통합 분석하기 위해 conv2D (convolution 2D, 2차원 합성곱) 기계학습 방법을 사용했고, 기계학습에는 LDAPS(Local Data Assimilation and Prediction System, 기상청 국지예보시스템) 수치모델자료로부터 풍향속, 연직속도 등 기상 데이터뿐 아니라 MODIS(Moderate Resolution Imaging Spectroradiometer) 위성으로부터 지형 환경 데이터도 추가하여 장기적으로 적용할 수 있도록 하였다. 3개 민간공항 데이터를 무작위로 4:1로 나누어 학습/테스트 데이터로 구분하고, 모델 수행 결과 학습 데이터로 생성한 모델이 테스트 데이터에 대해 분류 모델 평가 지표(Confusion Matrix) 중 정확도(accuracy)에 70~80%의 성능을 보였다. 본 연구를 통하여 개발된 급변풍 예측 시스템은 항공안전에 위협이 되는 급변풍 예측 정보를 제공할 수 있을 것이며, 이를 통해 공군 예보관이 항공작전에 미칠 영향과 위험을 조기에 판단하고 그로 인한 피해를 예방하는데 큰 기여를 할 수 있을 것이다.

주요어 : 급변풍, 공군, AI, 기계학습, 수치모델, MODIS
학 번 : 2020-26319

목 차

제 1 장 서 론 ····· 1

제 2 장 선행 연구 ····· 3
제 1 절 수치모델자료를 활용한 급변풍 예측 ····· 3
제 2 절 기계학습을 위한 데이터 처리법 ····· 4
제 3 절 환경에 따른 바람의 변화 ····· 8

제 3 장 연구 방법 ····· 10
제 1 절 연구 자료 ····· 10
1. 항공기상청 급변풍 경보 발표 자료 ····· 10
2. 기상청 국지예보모델(LDAPS) 자료 ····· 12
3. MODIS Land Cover Type 자료 ····· 15
제 2 절 데이터 처리 및 기계학습 ····· 17
1. 데이터 Re-sampling ····· 17
2. 데이터 feature 수 증가 ····· 20
3. 데이터 회전을 이용한 데이터 양 증가 ····· 21
4. Conv2D 기계학습 ····· 22

제 4 장 연구 결과 및 토의 ····· 26
제 1 절 모델 수행 결과 ····· 26
제 2 절 공군 비행단 적용 사례 ····· 30
제 3 절 토 의 ····· 34

제 5 장 결 론 ····· 35

참고문헌 ····· 36
Abstract ····· 41

표 목 차

[표 1-1] 항공기 기종별 중량과 최대속도 제원 ····· 2
[표 1-2] 최근 5년간 공항별 평균 운항 통계 ····· 2
[표 3-1] 공항 정보 ····· 10
[표 3-2] 양양공항 항공기상특보 ····· 11
[표 3-3] 급변풍 경보 기록 데이터 ····· 11
[표 3-4] 공항별 계기 접근 기준 거리와 고도 ····· 14
[표 3-5] 공항별 연간 급변풍 경보 유효시간 ····· 19
[표 3-6] 연도별 한반도 영향 태풍 현황 ····· 19
[표 3-7] 분류 성능 평가 지표 ····· 24
[표 3-8] 최종 학습 데이터 형태 ····· 25
[표 4-1] 민간공항 모델 수행 결과 ····· 26
[표 4-2] 급변풍 경보 발표 기록 일부 ····· 28
[표 4-3] 제주공항 모델 예측 사례 성능 지표 ····· 29

그 림 목 차

[그림 2-1] 제주공항 항공기상정보 ····· 5
[그림 2-2] 데이터 불균형(Imbalance) ····· 7
[그림 3-1] 양양공항 MODIS Land Cover 종류 자료 ····· 16
[그림 3-2] 폭풍일수 분포도 ····· 17
[그림 3-3] 회전 처리된 이미지 데이터 ····· 21
[그림 3-4] Conv2D 기계학습 방법 모식도 ····· 23
[그림 3-5] Adagrad optimizer 학습 방법 ····· 23
[그림 4-1] 제주공항 2020-11-21 연구 모델 시계열 자료 28
[그림 4-2] 제주공항 2020-11-22 연구 모델 시계열 자료 29
[그림 4-3] 청주공항 2019-03-02 연구 모델 시계열 자료 31
[그림 4-4] 청주공항 2019-03-02 기상청 LDAPS 시계열 · 31
[그림 4-5] 청주공항 2019-03-27 연구 모델 시계열 자료 32

제 1 장 서 론

항공기상청은 항공기와 비행장 시설 및 업무에 악영향을 미칠 수 있는 기상 현상에 대해 공항경보와 급변풍 경보를 발표하고 있다(항공기상청, 2021). 공항경보에는 태풍을 포함한 뇌우, 저시정 등의 기상현상이 포함되며, 지상 현상이 아닌 저고도 급변풍(Low Level Windshear)에 대해서는 급변풍 경보를 발표하고 있다. 급변풍은 방향 또는 속도가 급격히 변하는 바람으로(ICAO[1], 2005), 현재 항공기상청에서는 급변풍 탐지 장비를 활용하여 바람의 변화경향(Loss 또는 Gain)이 15kt(8m/s) 이상이 예상될 때 급변풍으로 정의(항공기상청, 2021)하고 경보를 발표하고 있다. 이러한 급격한 변화는 항공기의 비행 과정에 예상치 않은 힘을 발생하여 항공기 이착륙 시 자세 유지에 어려움을 주기 때문에, 항공기상 분야에서 안전 위협 요소로 간주되고 있다(Hon, K.K., 2019). 현재 민간공항에서는 급변풍 탐지 장비(LLWAS, Low Level Windshear Alert System)를 설치하여 급변풍 정보를 지원하고 있으나(항공기상청, 2021), 군에서 관리하는 공항에는 급변풍 탐지 장비가 설치되어 있지 않고 급변풍 정보도 제공되고 있지 않다. 전투기는 수송기나 여객기에 비해 속도는 2배에 달하여 이착륙 과정에서 속도의 변화폭이 크고, 민항기에 비해 작은 중량으로 인하여 상대적으로 급변풍에 취약하다[표 1-1]. 이처럼 물리적으로 급변풍에 취약함에도 상황에 따라서 시간을 조절하지 못하고 긴급 출격에 대비해야하는 군용기 조종사들의 안전 보장을 위해 급변풍 연구의 도입은 필수적이다.

2020년 9월, 항공기상청은 군 공항에 기 설치되어 있는 항공자동기상관측장비(AMOS, Automatic Meteorological Observation System)를 기반으로 군과 협의하여 급변풍 경보를 발표하기로 합의하였으나, 청주나 대구 비행단 등 군 기상대에서 지금까지 급변풍 경보에 대한 실질적인 지원을 받은 적은 전무하다. 청주와 대구공항은 연간 1만 편 이상의 항공편을 운항 중인데다[표 1-2], 공군 비행단은 코로나 백신 수송, 재외국민과 해외

1) 국제민간항공기구, International Civil Aviation Organization

파병부대 공수 등 긴급한 상황에서도 비행을 운용해야 하므로, 군용기의 안전 보장 및 독자적 군 작전 수행을 위해 군 공항 활주로의 급변풍 정보는 반드시 필요하다.

[표 1-1] 항공기 기종별 중량과 최대속도 제원

기종	자체중량	최대이륙중량	최대속도
F-5E	4.4t	11.2t	마하[2] 1.6
F-16C/D	8.6t	19.2t	마하 2.0
C-130H	34.4t	70.3t	마하 0.5
BOEING 747-8	19t	44t	마하 0.8

[표 1-2] 최근 5년간 공항별 평균 운항 통계[3]

공항명	운항(편)/년	공항명	운항(편)/년
인천	32만 8천	대구	2만 2천
제주	16만 4천	청주	1만 6천
김포	13만 7천	울산	6천
김해	9만 6천	여수	5천

그리고 현재 민간공항에서 운영 중인 LLWAS의 경우, 실시간으로 발생하고 있는 급변풍에 대한 정보만 제공 가능하고, 앞으로의 급변풍 발생에 대한 예측은 불가능하다(Golding, 2005). 따라서 일반적인 모든 공항에서 급변풍 정보에 대한 정보를 미리 제공받고 비행 계획에 반영하려면, 장비 설치와 별개로 급변풍 예측 정보를 생산하기 위한 독자적인 모델을 추가 개발해야 한다. 본 연구의 목표는 독자 개발 가능한 모델을 통해 비행단별로 비행을 계획하는 단계에서부터 해당 정보를 적극 활용하고, 사전에 급변풍을 피해 작전을 완수할 수 있도록 하는 것이다.

2) 음속에 대한 운동 물체의 속도의 비. 마하 1 = 시속 약 1,200km
3) 한국공항공사-정보공개-항공통계-공항별 통계

제 2 장 선행 연구

제 1 절 수치모델자료를 활용한 급변풍 예측

해외 기관에서는 고해상도 수치모델자료를 활용해 항공기 이착륙에 영향을 주는 급변풍 예측을 연구하고 있다. 홍콩기상청에서는 홍콩국제공항의 급변풍 예측을 위해 200m 해상도의 WRF(Weather Research and Forecasting) 수치모델자료를 활용, 공항에 이착륙하는 항공기의 3차원 이동 경로(glide path)를 따라 수치모델자료에서 가장 가까운(the nearest) 바람 벡터를 정풍 성분으로 분석하고, 그 변화를 바탕으로 급변풍 정보를 구현하였다(Hon, K.K., 2019). 하지만 현재 우리 기상청에서 사용 가능한 국지모델자료는 1.5km 해상도이고, 공군 비행단의 지리 정보 또한 세부적으로 접근할 수 없으므로 항공기의 경로까지 추적하기에는 어려움이 있다.

또한 홍콩기상청에서는 지형에 의한 급변풍 분포를 고해상도 AVM(Aviation Model) 모델을 사용해 구현하는 연구도 진행하였다. 모델자료를 통해 TDWR(Terminal Doppler Weather Radar) 레이다와 LIDAR(Light Detection and Ranging system) 시스템으로 생기는 사각(死角)을 없애고, 급변풍 정보를 급변풍이 발생하기 이전에 예보하기 위함인데, 비록 공군 비행단에는 TDWR 레이다나 라이다와 같은 장비가 모두 설치돼 있지 않지만, 수치모델로 정보를 미리 확인하고 극복할 수 있음을 확인할 수 있었다.

제 2 절 기계학습을 위한 데이터 처리법

본 연구의 목적은 다수의 공항에 통합적으로 적용 가능한 급변풍 예측 모델의 개발이다. 따라서 현재 급변풍 정보가 이용 가능한 민간공항 데이터를 기반으로, 공군 비행단에 통합 적용하는 모델을 개발하기 위해 기계학습을 적용하고자 한다. 항공기상청은 급변풍 예보를 포함한 공항기상정보를 웹사이트에 게시하고 있는데, 급변풍 예보는 [그림 2-1]과 같이 위험기상 예보에 급변풍 발생 가능성 시간과 개괄적인 설명으로 제공되고 있다. 현재 예보는 항공기상청에서 7개 민간공항에 대해서만 발표하고 있어, 군 공항에 대한 급변풍 정보를 추가로 생성하고, 공군 현역 기상예보관이 활용할 수 있도록 시간별 급변풍 정보를 생성하기 위해 기계학습을 적용하게 되었다. 기계학습을 적용하기에 앞서, 연구방법에 적용할 몇 가지 데이터 처리법에 대한 선행연구를 진행하였다. 이는 데이터 불균형 해소를 위한 데이터 처리법(under-sampling), 충분한 양의 데이터를 확보하기 위한 데이터 처리법(data rotation), 기계학습 효율을 높이기 위한 데이터의 feature 수 증가 처리법에 관한 내용이다.

항공기상정보 | 공항기상

항공 관련 기상정보를 안내합니다.

제주공항 제58호	
등록일	2021-09-29 06:00:00

- 오늘 낮부터 비 -

▶ 일기개황

○ **(오늘)** 현재(06시) 제주공항은 구름조금 끼어 있고 남동풍이 3kt(6km/h)로 불고 있으며, 시정은 10km입니다. 오늘은 서해상에서 동진하는 저기압의 영향을 받아 대체로 흐리고 낮(13시)부터 비가 내리겠습니다.

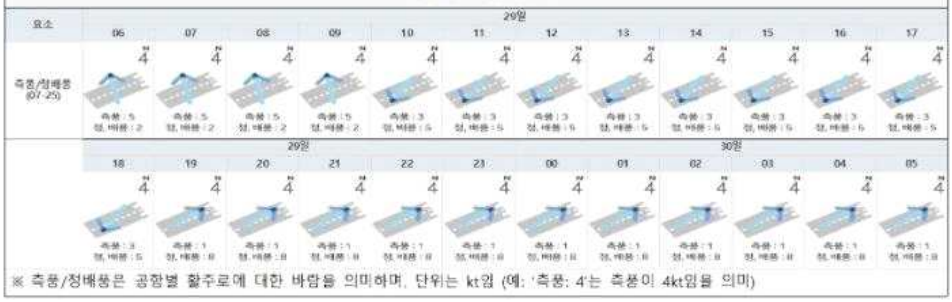
○ **(내일)** 오전까지 동해상에서 동진하는 저기압의 영향을 받다가 차차 벗어나 아침(9시)에 비가 그치고 흐리겠습니다.

최저/최고기온(℃)	낮 최고 체감기온(폭염영향수준)(℃)	예상 강수량(mm)
(오늘) -/27, (내일) 21/24	(오늘) 29, (내일) 25	15~20

▶ 위험기상예보

○ (급변풍) 오늘 오후(15시)부터 밤(23시) 사이 남풍류가 한라산을 돌아 들어오면서 풍향 변화에 의한 급변풍 발생 가능성이 있습니다.

시간별 측풍예보



※ 측풍/정배풍은 공항별 활주로에 대한 바람을 의미하며, 단위는 kt임 (예: '측풍: 4'는 측풍이 4kt임을 의미)

▶ 경보현황

○ 없음

[그림 2-1] 제주공항 항공기상정보[4)]

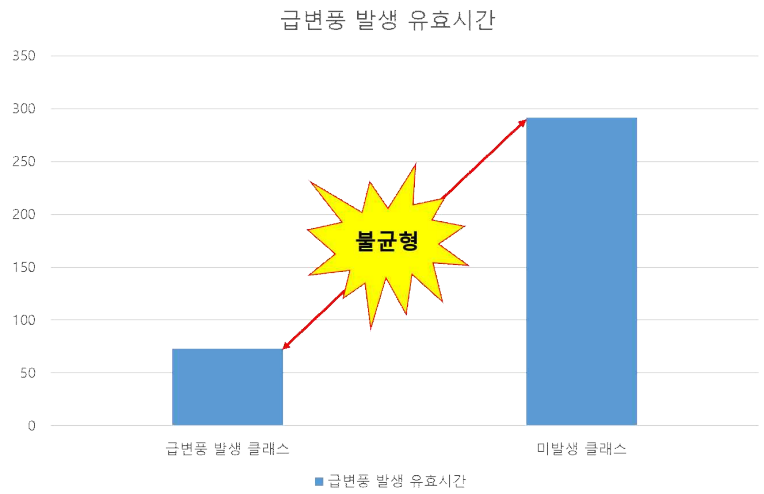
4) 항공기상청-항공기상-항공기상정보-공항기상정보

기계학습에서 분류하고자 하는 모든 데이터는 데이터 불균형(imbalance)이라는 특징을 가지고 있다(Liu, 2019). 본래 데이터 불균형은 클래스(class)[5]가 여러 개(multi-label)인 데이터의 학습에서 발생하는 고유의 특징인데, 불균형한 데이터의 학습은 특정 클래스에 편향된(dominant) 예측 결과로 이어진다. 이는 자연계에서 어떤 현상의 발생 여부를 예측하는 binary 모델의 경우에도, 미발생 클래스인 negative-dominant 편향이 나타나, data imbalance를 고려해야한다(Estabrooks, 2004).

예를 들어, 상층운, 중층운, 하층운 사진을 학습하여 구름의 고도를 3가지로 구분하는 모델에, 상층운 사진 100장, 중층운 사진 50장, 하층운 사진 10장을 학습하면, 모델은 상층운에 편향된 예측 결과를 생산하게 되고, 이런 학습 데이터의 구성을 '불균형하다'고 한다. '균형' 잡힌 현상은 위와 같은 구름 클래스의 경우 각 고도별로 균등하게 발생하거나, 강수의 유무 같은 경우에는 강수 확률이 50%인 경우를 말할 수 있겠다.

본 연구는 자연계 현상의 발생 여부에 초점을 두기 때문에 '비대칭적이다'고 볼 수도 있는데, 급변풍의 경우 시간 단위로 분석했을 때, [표 3-5]와 같이 1년 중 20% 미만의 데이터만이 급변풍 발생시간으로 분석된다. 즉, 자연계 데이터를 그대로 학습한다면, 일 단위로 환산시 급변풍 발생 클래스 데이터는 약 70일, 급변풍이 발생하지 않은 클래스 데이터는 약 295일로 구성된다. 70일과 295일 간 클래스 크기의 불균형(비대칭) [그림 2-2]으로 인해 모델은 급변풍이 발생하지 않은 클래스에 편향된 예측 결과를 생산하게 될 것이다.

5) 데이터를 분류하는 종류. 머신러닝에 통과 후 도출되기를 기대하는 정답을 일컫는 label(레이블)과 혼용하기도 한다.



[그림 2-2] 데이터 불균형(Imbalance)

불균형한 데이터로의 학습은 더 많은 양을 가진 데이터 클래스, 폭풍일수의 경우에는 폭풍이 발생하지 않은 클래스에 편향된 예측 결과로 나타나게 되므로(Estabrooks, 2004), 데이터 불균형 해소를 위한 처리 과정은 급변풍 예측 성능 향상을 위해 필수적이다. 이런 과정은 모델에 데이터를 입력하기 전에 처리하는 전처리 방법과 모델이 수행된 후 처리하는 후처리 방법이 있고, 전처리 방법 중에는 무작위 샘플링(random re-sampling)[6] 방법이 있다(Branco, 2015). Multi-label 데이터의 경우엔 데이터 클래스가 더 많아지다 보면 class 간 불균형의 정도가 더욱 심해지는데, 이렇게 불균형이 심할 때는 random under sampling 방법을 적용해 볼 수 있다(Liu, 2019). 급변풍의 경우에도 폭풍일수 통계와 같이 불균형이 매우 심하므로[표 3-5], random under sampling 방법을 적용하였다. 데이터 수준에서 직접 데이터의 불균형을 처리하는 방법 외에, 기계학습 과정에서 loss 함수(손실 함수)를 수정하여 기계학습의 갱신 과정에서 데이터 불균형이 제어될 수 있도록 하는 알고리즘 수준의 방법도 있다(Johnson et al., 2019).

6) Table 3: Pre-processing strategy types

Under sampling 방법을 적용하면, 전체 데이터에서 무작위로 데이터를 제외하기 때문에, 학습 데이터의 양이 줄어드는 단점이 있다. 충분한 양의 학습 데이터 확보를 위해 기존 데이터의 형태를 조금씩 바꾸어 반복 학습하는 방법으로 보완할 수 있는데, 특히 이미지와 같이 2차원의 형태로 이루어진 데이터의 경우, 각각의 데이터를 회전하여 학습하는 방법이 있다. 구체적으로, 0도, 90도, 180도, 270도 4가지 방향에 대한 회전을 적용했을 때 성능이 가장 크게 좋아졌다(Gidaris, 2018). 또한 데이터 양의 증가보다는 데이터 feature[7] 수의 증가가 모델 성능 향상에 효과를 보이며(고경규, 2021), 지역기후모델모의 연구에서는 지상 풍속의 변화를 모의하는데 해면기압(MSLP, Mean Sea Level Pressure)이 상관관계를 가진다고 제시하였다(Jeong, Dae Il, 2019). 따라서 바람 데이터 외에 기압 등 다른 요소를 추가해보게 되었다.

제 3 절 환경에 따른 바람의 변화

바람의 변화는 지형과 기후 등 환경에도 영향을 받는다. 항공기 이착륙에 영향을 주는 급변풍은 1,600ft(약 500m) 이하 저고도 상층 및 지상풍의 영향을 받는데(ICAO, 2005), 기후변화 시나리오 RCP2.6, RCP4.5, RCP8.5에 따라 한반도의 연평균 상층풍을 비교해보면, 2020년부터 2040년 사이에 2.4~3.8% 가량 풍속이 증가함을 제시하였다(김선애, 2015).

일반적으로 급변풍은 지형이 복잡하거나, 고층 건물이 많거나, 혹은 산지/해안 지형에서 강한 바람이 발생하거나 풍향이 바뀌기 쉬운 조건에 많이 발생하는 것으로 알려져 있다. 또한 계절상으로는 대부분의 지역에서 봄철에 가장 많은 발생빈도(35~64%)를 보였다(Lin, C. et al., 2021). 반면 지상풍의 경우 지형(Land cover) 변화에 영향을 받는데, 도시 발전에 따라 roughness length가 큰 urban 지역이 증가할수록, 지상 풍속은 감소하는

7) 특징. 학습 모델로 정답을 도출하기 위해 고려하는 데이터의 특징을 일컬으며, attribute(속성)과 혼용하기도 한다.

경향을 보였다(Fu, Donglei, 2020; Li, Yupeng, 2019). 기후변화로 상층풍속은 증가하는데, 도시화로 지상 풍속은 감소한다면, 결과적으로 상하층 풍속 차이의 심화로 급변풍 발생 위험이 증가하게 된다. Land cover 종류가 'urban and built-up'과 같이 도시화한 종류로 많이 변할수록, 고층 건물이 많아지거나 지면 마찰이 증가하기 때문에, 지상 풍속 관측은 감소하고, 건물 꼭대기에서의 상층풍속 관측은 강풍이 많이 발생하게 되어 급변풍이 발생하는 쉬운 조건이 형성된다. 키가 큰 나무가 많거나 고도가 높은 산지에서도 같은 원리로 급변풍이 발생하기 쉬워지며, 해안에서는 오히려 지면 마찰이 낮아져 상층에 강한 풍속의 바람이 저고도까지 혼합(mixing)되어 급변풍이 발생하기 쉬워진다. 공군 비행단 지역의 도시화가 진행되면 급변풍 경보에 대한 위험이 증가할 것이므로, 환경에 따른 효과가 모델 개발 과정에 포함돼야 할 것으로 판단하였다.

제 3 장 연구 방법

제 1 절 연구 자료

1. 항공기상청 급변풍 경보 발표 자료

현재 민간공항에 대해서는 항공기상청이 2006년 11월[8]부터 급변풍 경보를 발표하고 있다(항공기상청, 2021; 항공기상청, 2015). 실시간 급변풍 등 공항특보 발표 현황은 항공기상청 웹사이트[9]에서 확인 가능하며, 과거 5년(2016~2020년)에 대한 급변풍 경보 발표 자료는 항공기상청에 요청하여 제공받았다. 대상 공항은 항공기상청이 급변풍 경보를 제공하고 있는 인천, 제주, 양양공항 등이다.

[표 3-1] 공항 정보

공항명	WMO ID	위도	경도	활주로 방향
인천공항	47113	37.4625	126.4392	330
김포공항	47110	37.5683	126.7756	320
제주공항	47182	33.5167	126.5000	250
양양공항	47092	38.0667	128.6667	330
김해공항	47153	35.1692	128.9344	360
대구공항	47142	35.8908	128.6611	310
청주공항	47128	36.7167	127.5000	240
울산공항	47151	35.5933	129.3522	360
여수공항	47167	34.8467	127.6125	350
광주공항	47158	35.1250	126.8117	220
사천공항	47161	35.0833	128.0666	240

공항의 위경도 좌표는 기상청 방재기상정보시스템[10]의 AWS 지점정보를, 활주로 방향은 구글어스(Google Earth) 영상을 통해 확인했다.

---

8) 인천공항 급변풍경보 발표 시작 기준
9) amo.kma.go.kr
10) afso.kma.go.kr

급변풍 경보 발표 자료는 공항별로 언제 발표했고(발표시간), 언제부터(발효시간) 언제까지 유효한지(해제시간) 기록하고 있다. 예를 들면, 양양공항의 2018년 1월 1일 기록을 보면, [표 3-2]와 같은 정보를 확인할 수 있다.

[표 3-2] 양양공항 항공기상특보

[항공기상특보] 양양공항 - 발표시간 : 2018-01-01 23:47 - 발효시간 : 012347 - 해제시간 : 021000

시간은 모두 UTC[11] 시간을 사용하며, 발효/해제시간은 발표시간으로부터 일/시/분 순이다. 이 기록에서 해제시간으로부터 발효시간을 뺀 시간 동안 유효했던 급변풍 경보를 시간별로 정리하여 급변풍 경보 기록 데이터를 생성하였다. 상기 양양공항의 경우, 급변풍 경보가 2018년 1월 1일 23시 47분부터 2일 10시까지 발표되었으므로, 2일 0시부터 10시까지의 기록 데이터에는 급변풍 경보가 발표되었음을 의미하는 숫자 1이, 그 외의 시간에 대해서는 숫자 0이 저장되도록 하였다.

[표 3-3] 급변풍 경보 기록 데이터

공항 ID	연	월	일	시(UTC)	경보 여부
47092	2018	1	1	23	0
47092	2018	1	2	0	1
47092	2018	1	2	1	1
47092	2018	1	2	...	...
47092	2018	1	2	10	1
47092	2018	1	2	11	0

11) UTC(Coordinated Universal Time) : 협정세계시. 그리니치 표준시(GMT)와 같으며, Z(Zulu, 줄루) time과 같다. 한국표준시(Korean Standard Time)는 UTC보다 9시간 빠르다(+9).

2. 기상청 국지예보모델(LDAPS) 자료

현재 기상청 기상자료개방포털[12]을 통해 제공되고 있는 수치모델자료 중 LDAPS(Local Data Assimilation and Prediction System) 자료는 단기 국지예보모델로, 전지구모델로부터 초기 경계장을 3시간 간격으로 제공받아 일 8회 생산되고 있다. 공간해상도는 1.5km, 연직으로는 약 40km까지의 고도를 70층으로 나누어, 저고도에서는 25hPa 간격의 등압면 자료를 사용할 수 있다. 예측시간은 매 초기시간으로부터 48시간까지이다. LDAPS의 풍속 예측 성능은 AWS 설치 지점의 지형에 관계없이 대체로 1m/s 정도 과대모의하며, 지형에 따라서는 지방의 평지(Rf, Rural flat) 지형에서 오차가 가장 작았다. 그리고 LDAPS 내에 입력된 고도가 영향을 주는데, 지방의 산지(Rm, Rural mountainous)의 경우 LDAPS가 고도가 실제고도보다 낮을수록 풍속을 과소모의하였다(Kim, Dong-Ju et al., 2020).

일반적으로 자연계의 운동은 비선형적인 운동으로 수치근사를 사용해 예측하기 때문에, 기상 예측은 초기값에 의한 오차를 가지고 있다[13] (Holton, 2004). 따라서 연구 대상인 5년(2016~2020년) 중 학습에 사용할 자료는 3시간 간격인 매 initial time에서 가장 가까운 h00~02(초기~2시간 예상장) 자료만 모아서 구성했고, 검증에 사용할 자료는 공군 기상대의 오전 시간 작전 판단시점을 고려하여 오전에 제공되는 12UTC 자료의 h03~h27(3~27시간 예상장) 자료로 다음날 24시간 동안을 분석하는 데 중점을 두었다.

LDAPS 등압면 자료에는 850hPa 이하 저고도에 대해서는 25hPa 간격으로 데이터가 저장되어있다. 표준대기고도와 급변풍 영향 고도 1,600ft (약 500m)를 고려하여, 950hPa(약 530m, 1,700ft) 이하의 고도에 대한 데이터를 적용하였다. 연구 대상 고도는 950hPa, 975hPa, 1000hPa 3개 고도이다.

---

12) data.kma.go.kr
13) Section 13.3 Numerical Approximation of the Equations of Motion

LDAPS에는 기온, 바람 등 여러 가지 기상정보가 고도별로 저장되어 있다. 바람 정보로는, 수평 바람 성분을 나타내는 U component of wind, V component of wind, 그리고 연직 운동을 나타내는 Vertical Velocity가 있다. 수평 성분 바람의 경우, 동일한 방향에 대한 풍속 차이에 따라 와도(渦度)가 발생하므로[14](Holton, 2004), 와도의 발생에 따른 연직 운동의 강도를 분석하기 위해 기준 풍향을 정할 필요가 있었다.

항공기가 저고도에서 접근할 때 활주로와 나란한 방향으로 접근한다[15]는 점에서, 활주로 방향을 기준으로 한 정풍 성분의 풍속 변화를 분석하였다. U, V component of wind 벡터 값을 합성하여 분석할 각 공항의 활주로 방향에 따라 정풍성분 풍속으로 분석하였다. 바람 정보 외에도 학습 데이터(Train data[16])의 feature를 늘리기 위해 다른 종류의 데이터도 포함시켰는데, LDAPS는 상층 데이터를 사용하기 때문에 기압 정보인 MSLP를 사용(Jeong, Dae Il, 2019)할 수 없어, Geopotential Height[17] 데이터를 포함하였다. 결과적으로, 수평 정풍성분 풍속, Vertical Velocity, Geopotential Height 3가지 데이터를 사용할 수 있었다.

급변풍 경보는 활주로에 근접했을 때 영향을 주는 바람에 대한 경보이므로, 사용할 LDAPS 950hPa(약 1,700ft) 고도에 대한 항공기의 수평 범위를 정해야 한다. 국내공항은 국토교통부 항공정보서비스(AIS, Aeronautical Information Services)에 항공기 접근 절차를 고시하고 있는데, 공항별로 약간의 차이는 있으나 대체로 5NM[18](9km) 거리에 접근할 때 약 1,800ft(550m) 고도로 접근하도록 안내하고 있다. 따라서 LDAPS로부터 데이터를 추출할 수평 범위는 반경 9km로 판단하였고, LDAPS의 수평 해상도 1.5km를 고려하면, 공항에 가장 가까운 LDAPS 위경도 좌표 기준 144개[19]의 데이터가 포함되도록 하였다.

---

14) Section 4.2 Vorticity
15) 국토교통부 항공정보서비스
16) 정답(label, 레이블)이 부여된 특징 벡터의 집합
17) 지위고도. 지면에서 특정기압이 되는 높이로, 고기압일수록 높게 나타난다.
18) Nautical Miles, 1NM = 약 1,800m
19) {2*(9km/1.5km)}2=144

[표 3-4] 공항별 계기 접근 기준 거리와 고도

공항명	계기 접근 기준		
	거리	고도	고도/5NM [20]
인천	5NM	1,790ft	1,790ft
김포	4NM	1,400ft	1,750ft
제주	5NM	1,800ft	1,800ft
양양	4.5NM	1,700ft	1,890ft
김해	6.4NM	2,100ft	1,640ft
대구	7.2NM	2,700ft	1,875ft

6개 공항 5NM 계기 접근시 평균 기준 고도 = 1,790ft(550m)

20) 활주로로부터의 거리를 5NM로 선형환산시 접근 고도

3. MODIS Land Cover Type 자료

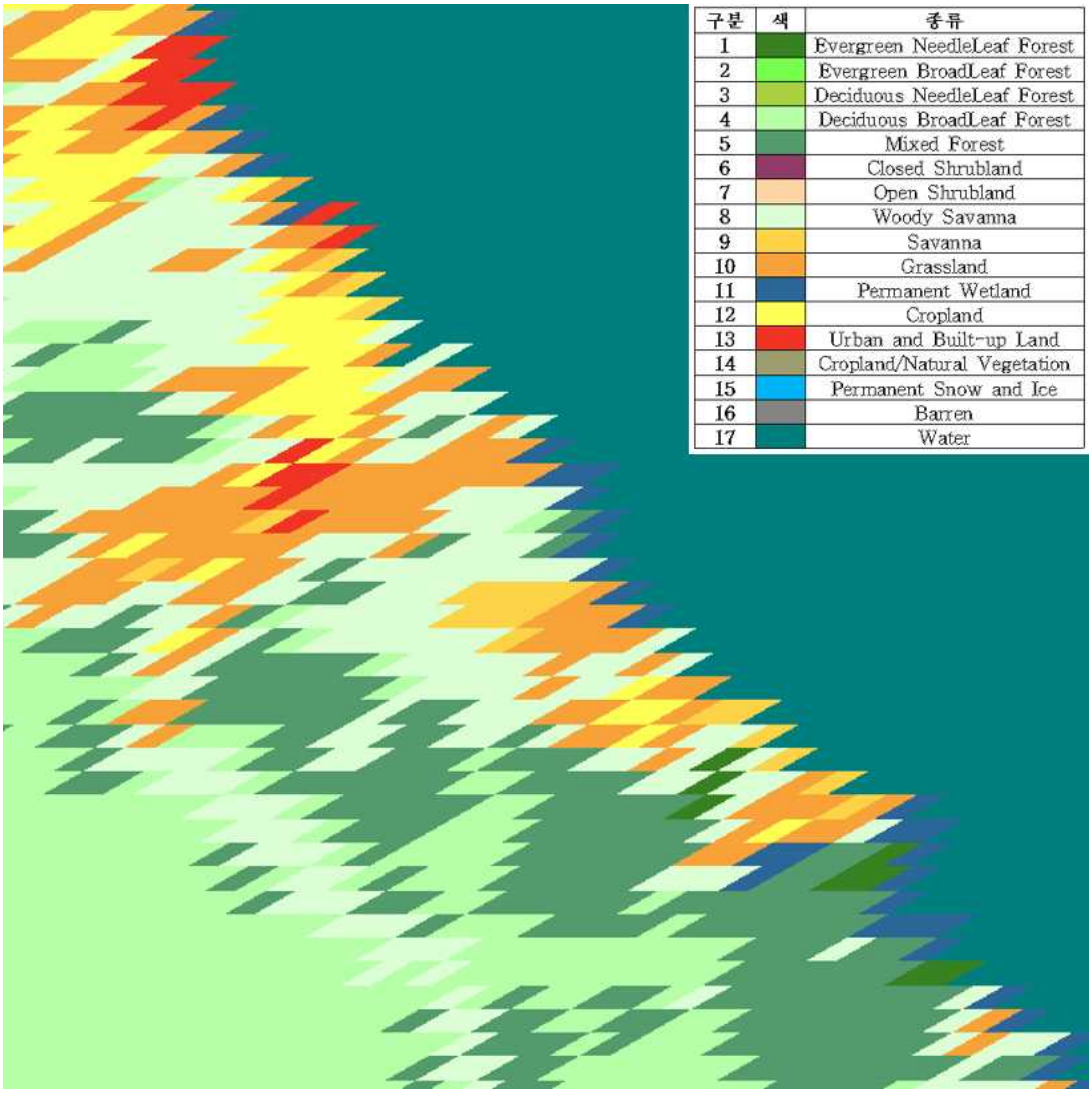
MODIS(Moderate Resolution Imaging Spectroradiometer)는 미국 NASA(美항공우주국, National Aeronautics and Space Administration)의 Terra/Aqua 위성에 탑재된 장비로, MODIS 산출물 중 MCD12Q1 자료는 연간 Land Cover Type을 500m 해상도로 제공[21]하고 있다.

기상 데이터 외의 지형 영향을 반영하기 위해 MODIS Land Cover Type 자료를 포함시켰다. 최초 모델을 구성할 때, 최대한 관측에 가까운 자료로부터 예측 결과를 얻기 위해 수치모델자료가 아닌 위성자료로부터 Land Cover 자료를 수집하였다. 이 때, MODIS 위성자료는 500m 해상도인데 LDAPS 자료는 1.5km 해상도로 차이가 있어, 기계학습을 위해 해상도를 맞출 필요가 있다. LDAPS 자료에서 추출한 범위와 동일한 범위의 MODIS Land Cover 자료를 이미지화한 뒤, 12x12 격자별로 분석하여 각 격자에서 가장 많은 빈도를 차지하는 Land Cover Type이 그 격자를 대표하는 12x12 배열 형태의 데이터를 생성하였다. [그림 3-1]은 MODIS Land Cover 자료를 양양공항의 위치를 기준으로, LDAPS 자료의 범위와 같은 범위만큼 이미지화한 결과이다. [그림 3-1]에 나타난 색은 우측 상단 표기와 같은 Land Cover 종류를 나타내며, 각각 다른 숫자로 구분된다.

이 방법을 통해 공항별로 LDAPS 데이터와 MODIS Land Cover 데이터가 같은 스케일(12x12)로 조정되었고, input 데이터에는 land cover type feature를 추가할 수 있게 되었다. MODIS 자료는 연도별로 제공되므로, LDAPS 자료의 연도와 맞추어 같은 연도에는 동일한 land cover 데이터가 결합되도록 하였다.

---

21) NASA, Earthdata



[그림 3-1] 양양공항 MODIS Land Cover 종류 자료[22)]

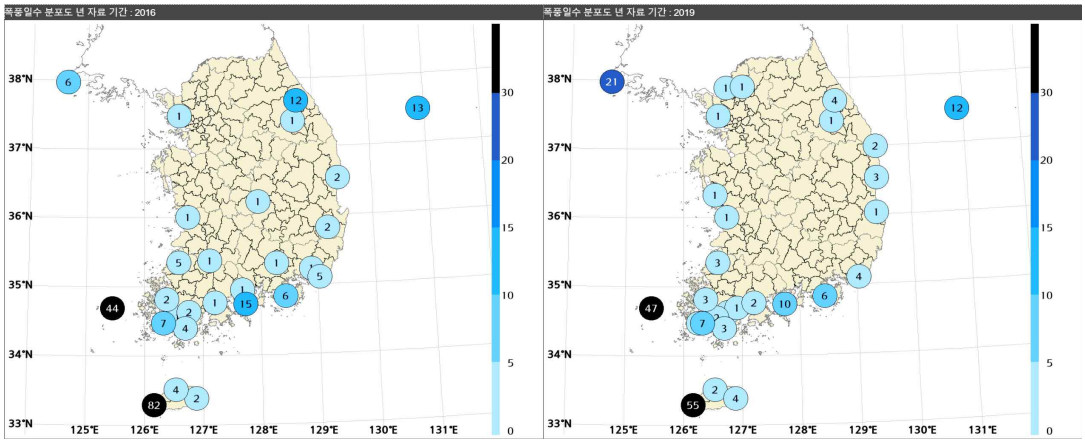
위 이미지의 중심은 양양공항에서 가장 가까운 LDAPS 상의 격자이다. 북동쪽에 물을 나타내는 Land Cover Type부터, 좌상단(북서쪽)부터 우하단(남동쪽)을 가로지르는 동해안 해안선, 그리고 남서쪽 내륙으로는 삼림을 보여주고 있다.

22) 2019년 MODIS 자료

제 2 절 데이터 처리 및 기계학습

1. 데이터 Re-sampling

현재 제공되고 있는 급변풍 데이터는 시기와 지역에 대해 불균형(imbalance)하다. 시간적으로는 폭풍일수와 같이 발생하지 않은 날이 발생한 날보다 많으며, 지형과 연도에 따라서도 불균형을 보였다. 지형에 따라서는 폭풍일수 분포도를 살펴보면, 지형에 의한 마찰이 적은 도서지역과 해안과 직접 인접한 남·동해안 지역에는 더 많은 폭풍이, 상대적으로 주변 섬 또는 지형에 둘러싸이거나 내륙에는 더 적은 급변풍이 발생하였다.



[그림 3-2] 폭풍일수 분포도[23] (좌측 2016년, 우측 2019년)

[그림 3-2]는 연도와 지형에 따른 불균형을 보여주고 있다. 2016년 폭풍일수 분포도(좌측)에서 제주와 흑산도지역이 가장 많은 분포를 보이고, 그 다음 울릉도지역과 사천, 양양 일대가 많은 분포를 보인다. 전라 남서해안 지역은 상대적으로 적은 분포를 보이는데, 이는 제주와 흑산도와 같은 섬 지형이 직접적인 해상영향으로부터 막아주고 있어, 남·동해안에서

23) 기상자료개방포털, [표 3-5]에서 제주공항이 가장 많은 급변풍 경보 유효시간을 보여, 제주공항의 최대-최소 급변풍 기록을 보인 2016-2019년 폭풍일수를 비교하였다.

더 많은 분포를 보이는 것으로 분석된다. 2019년 폭풍일수 분포도(우측)에서도 제주, 흑산도, 백령도, 울릉도와 같은 도서지역은 여전히 다른 지역에 비해 많은 분포를 보이고 있으나, 제주지역 분포가 크게 감소하고 백령도 지역이 크게 증가한 점, 내륙에서 10회를 초과한 분포를 보이지 않은 점은 2016년과 차이를 보이고 있다.

이런 데이터 불균형은 급변풍 경보 발표 기록에서도 동일하게 나타났는데, 태풍기간을 제외한 1년 약 6,500여 시간 중 대부분의 공항에서는 연간 1,000시간 이하의 급변풍 경보가 발표되었다. 하계 태풍기간을 제외한 이유는 하계에는 장마로 인한 호우와 뇌우, 태풍이 도래하면 강풍 등의 사유로 급변풍 외의 요인으로 항공작전이 제한되기 때문이다. 이렇게 제외한 기간은 연도별 태풍 영향 시기를 조사하여 [표 3-6]과 같다.

하계를 제외한 기간 내에서 전 지점에 대해 시간상으로 평균 15% 정도의 시간동안만 급변풍 경보가 유효했으며, 나머지 85%의 시간동안에는 급변풍이 발생하지 않아 negative-dominant 데이터가 된다. 또한 2016년 제주공항에 발표된 급변풍 경보는 1,061시간에 달하는 반면, 2019년 인천공항에 발표된 급변풍 경보는 347시간에 불과했는데, 이는 지점마다 지역적인 편차가 3배 이상 나타나고 있음을 의미한다. 이러한 지역이나 연도에 따른 데이터 불균형 문제를 해결하기 위해, 공항과 연도별로 각각의 데이터에 대해 무작위 under-sampling 작업을 수행하였다.

데이터 under-sampling은 모집단에서 더 적은 숫자의 표본을 추출하여 데이터를 구성하는 것으로, [표 3-5]와 같이 공항과 연도에 따라 급변풍 경보 유효시간의 차이가 큰 경우에는 under-sampling 방법으로 공항과 연도에 따른 편향을 줄일 수 있다. 본 연구에서 under-sampling 방법은 공항과 연도별로 당시에 발표되었던 급변풍 경보 유효시간에 따라, 각 데이터의 미발표 시간 중에서 유효시간의 양만큼 데이터를 무작위로 추출했기 때문에, 특정 연도 또는 공항에서 특히 급변풍이 적게 부는 환경이었더라도 다른 조건과 같은 비중으로 모델에 학습될 수 있을 것이다.

[표 3-5] 공항별 연간 급변풍 경보 유효시간 (1~6월, 10~12월 間)

공항명	급변풍 경보 유효시간				총 유효시간
	2016년	2017년	2018년	2019년	
인천공항	368	410	353	347	1,478
제주공항	1061	1033	793	684	3,571
양양공항	760	419	441	423	2,043
김포공항	73	83	156	149	461
여수공항	298	150미만	89	200미만	750미만
울산공항	150미만	100미만	200미만	100미만	550미만

[표 3-6] 연도별 한반도 영향 태풍 현황[24)]

연 도	최초발생(태풍명)	최종종료(태풍명)	총 기간(갯수)
2016	9.13. (제16호 말라카스)	10.6. (제18호 자바)	9.13.~10.6.(2개)
2017	7.2. (제3호 난마돌)	9.18. (제18호 탈림)	7.2.~9.18.(3개)
2018	7.4. (제7호 쁘라삐룬)	10.7. (제25호 콩레이)	7.4.~10.7.(5개)
2019	7.16. (제5호 다나스)	10.3. (제18호 미탁)	7.16.~10.3.(7개)
2020	8.9. (제5호 장미)	9.7. (제10호 하이선)	8.9.~9.7.(4개)

연도별 한반도 영향 태풍 기간을 조사하여, 태풍 기간을 제외하고 급변풍 경보 기록 데이터를 생성하였는데, 7~9월은 태풍기간으로 판단하여 연구 대상기간에서 제외하였다(기상청, 2011). [표 3-6]에서 마지막 태풍종료일에 따라 10월 일부기간도 제외하였다.

24) 기상청 방재기상정보시스템-태풍정보-태풍목록.
2018년 제7호 태풍 쁘라삐룬은 7월 이전에 발생했으나, 7. 1. 03시 기준 일본 오키나와 남쪽 약 390km 부근 해상에 위치했고, 당시 강풍반경은 250km로 분석돼(기상청, 2018, 「태풍정보 제7-8호」), 한반도 영향은 7월 이후로 판단하였다.

2. 데이터 feature 수 증가

연직으로 인접한 두 층에서의 바람 벡터 차이는 온도풍의 형태로 말할 수 있고[25](Holton, 2004), 온도풍과 온도이류는 연직 운동을 유발하기 때문에[26](Holton, 2004), 연직 운동 영향을 반영하기 위해 Vertical Velocity 외 추가적으로 인접한 등압면의 수평 풍속 차이를 고려해 볼 수 있다. 즉 950-975hPa과 975-1000hPa 간 수평 풍속 차이 feature를 추가하였다. 따라서, 950hPa, 975hPa, 1000hPa 3개 고도에 대한 정풍 성분 수평 풍속, Vertical Velocity, Geopotential Height(3x3=9가지), 연간 MODIS Land Cover Type(1가지), 그리고 2개 인접한 고도 간 수평 풍속 차이(2가지)까지 총 12가지 feature의 데이터를 생성하였다.

25) Section 3.4 The Thermal Wind
26) Section 6.2 The Quasi-geostrophic Approximation

3. 데이터 회전을 이용한 데이터 양 증가

연간 급변풍 경보 유효시간에 따라, 공항과 연도별로 1년 약 6,500여 시간 중 평균적으로 600시간 내외의 유효시간이 분석되었다[표 3-5]. 학습에 사용한 3개년(2016~2018년), 3개 공항(인천, 제주, 양양)을 통합하여도, 학습 가능한 유효시간은 약 5,000시간에 불과하였다. 기상 분야에서 효율적인 기계학습을 위해서 10,000시간 이상의 데이터를 학습한 사례(Hon, C.C., 2020)가 있으므로, 같은 양 또는 그 이상의 학습 데이터의 양을 확보할 필요가 있었다. 이미지 데이터 회전(Image-rotation) 방법은 이미지 분류를 위한 기계학습을 시행할 때 사용하는 방법으로, 본래 학습할 이미지를 회전하여 다시 학습하게 하는 방법이다[그림 3-3]. 이는 다양한 각도를 적용할 수 있으며, 특히 레이블 된 데이터를 구하기 어려운 경우, 쉽게 학습 데이터의 양을 늘리고 feature를 학습할 수 있어, 학습 효율을 증가시키는 데 사용된다. Gidaris(2018)는 4가지 방향(0도, 90도, 180도, 270도)으로 회전한 데이터로의 학습이 가장 좋은 효율을 보인다고 제시하여, 본 연구에서도 sampling된 데이터를 4가지 방향으로 회전/통합하여 충분한 데이터의 양을 확보하였다.

[그림 3-3] 회전 처리된 이미지 데이터[27)]

위와 같이 정사각형 모양의 배열 데이터를 회전 처리한 후, 모두 통합하여 학습 데이터의 양을 늘릴 수 있었다.

---

27) Gidaris(2018), Figure 1

4. Conv2D 기계학습

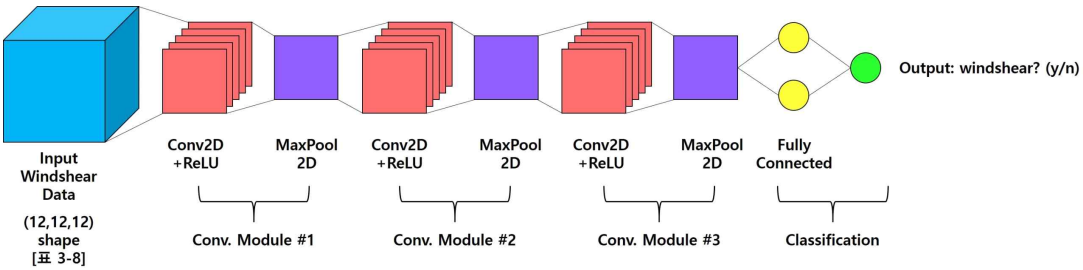
확대된 최종데이터를 사용하여 Conv2D(Convolution 2D, 2차원 합성곱) 기계학습을 진행하였다. Conv2D는 이미지와 같이 여러 feature로 이루어진 데이터의 특징을 학습하여 분류하는데 사용되는 기계학습 방법이다[그림 3-4]. Conv2D와 같은 현대 CNN(Convolutional Neural Network, 합성곱 신경망) 기계학습은 1990년대 mnist(Modified National Institute of Standards and Technology) 손글씨 데이터베이스를 사용하여 글씨 인식을 학습하는데 처음 사용되었다(Yann LeCun et al., 1998). 이후, 보다 복잡한 이미지로부터 feature를 학습해서 인식하는 기술이 개발되고 있다[28]. 본 연구는 급변풍 여부를 인식하기 위한 학습데이터를 이미지와 같은 2차원 구조를 갖도록 설계하여 학습시키고자 한다.

기계학습을 수행하는 과정에서, 모델 성능 결과에 따라 성능이 좋은 모델을 저장하면서 학습을 수행하도록 설정했는데, 이 때 성능 결과를 비교할 수 있도록 테스트 데이터를 분리할 필요가 있었다. 따라서 전체 데이터 중 무작위 80%를 학습 데이터(Train data)로, 20%를 테스트 데이터(Test data[29])로 분리하여, 기계학습을 수행하였다. 최종적으로 가장 좋은 성능을 보인 모델은 검증 데이터를 실험하여 테스트 결과와 비교하였다. 검증 데이터(Validation data)는, 본 연구가 급변풍 탐지 장비가 설치되어 있지 않은 공군 비행단에 적용할 목적이므로, 학습 데이터에 사용되지 않은 제3의 공항, 그리고 2019년 데이터[30]를 추가 사용하였다. 검증 데이터는 예측 결과를 예단할 수 없으므로, under-sampling 없이 사용하였다.

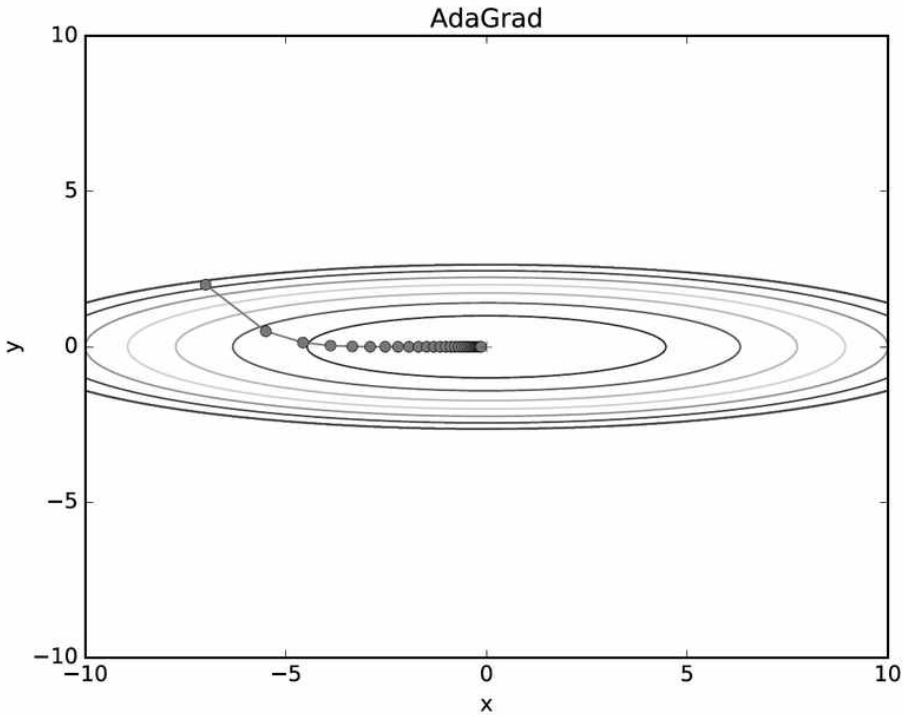
---

28) The History of Convolutional Neural Networks
29) 정답(레이블)을 아는 특징벡터의 집합. 학습된 모델에 이 집합을 통과시켜 정답이 도출되는지 확인하여 모델 성능을 평가한다.
30) 2020년 MODIS 자료는 2022년 6월 업데이트 예정(NASA)

기계학습은 3층의 conv2D module을 통해 수행하였다. [그림 3-4]는 본 모델에서 수행되는 conv2D 기계학습 방법으로, 총 3개의 conv. module을 통해 feature를 학습하고, windshear 발생여부를 (y/n)로 판단하는 모식도이다. 본 연구 데이터는 (12, 12, 12) 모양의 데이터로 처리할 양이 많아, 학습의 효율성을 고려하여 3개 모듈로 진행하였다.



[그림 3-4] Conv2D 기계학습 방법 모식도



[그림 3-5] Adagrad optimizer 학습 방법[31)]

31) Deep Learning - 학습관련 기술

기계학습은 [그림 3-5]와 같이 loss 함수(손실 함수)를 통해 loss를 구하고[32], 학습 과정에서 loss를 줄여가는 방법(optimizer)에 따라 수행된다. 본 연구에서 loss 함수로는 torch 패키지의 CrossEntropyLoss를, optimizer로는 Adagrad를 사용하였다. 최종 모델은 기계학습 수행 간 테스트 결과에서 f1-score[33] 가 가장 좋은 모델을 채택하였다.

[표 3-7] 분류 성능 평가 지표

구분		예상	
		Positive	Negative
실제	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

positive = 급변풍이 발생한 경우

accuracy = (TP + TN) / total

recall[34] = TP / (TP + FN)

precision[35] = TP / (TP + FP)

f1 score = 2 * precision * recall / (precision + recall)

32) [그림 3-5] y축

33) F1 score. 재현율을 높이면 무조건 맞추기 위해 과대모의하여 정밀도가 낮아질 수 있고, 정밀도를 높이면 정답일 확률이 높은 같은 것만 맞추기 때문에 재현율이 낮아질 수 있다. 재현율과 정밀도의 복합적인 성능을 평가하기 위해, f1 score는 재현율과 정밀도의 조화 평균으로 산출한다(Sasaki, 2007).

34) Recall(재현율). 실제 참 값 중 모델이 얼마나 놓치지 않고 많은 참 값을 예측해 내는지 평가하는 지표. 예를 들어, 실제 급변풍이 분 날 중 며칠의 급변풍을 잘 예측해 냈는지 평가할 수 있다(Makhoul, 1999).

35) Precision(정밀도). 모델이 예측한 값 중 얼마나 많이 참 값에 적중했는지 평가하는 지표. 예를 들어, 모델이 급변풍 발생일로 예측한 날 중 며칠동안 실제로 불었는지 평가할 수 있다. 모델이 급변풍 발생을 과대모의하는 것을 평가할 수 있다.

[표 3-8] 최종 학습 데이터 형태

	구분					레이블	데이터				
	A	B	C	D	E	F	G	H	I	J	K
1											
2											
...											

A : 공항 WMO ID

B : 연(Year)

C : 월(Month)

D : 일(Date)

E : 시(UTC)

F : 급변풍 경보 발표 여부

* A~F행 : [표 3-3] 급변풍 경보 기록 데이터와 동일

G : 고도별 정풍 성분 풍속 (3고도, 144개)

H : 고도별 Vertical Velocity (3고도, 144개)

I : 고도별 Geopotential Height (3고도, 144개)

J : Land Cover (1고도, 144개)

K : 인접한 등압면의 정풍 성분 풍속 차 (2고도, 144개)

* G~K행 : 총 12 features, (12,12,12) shape

제 4 장 연구 결과 및 토의

제 1 절 모델 수행 결과

학습 데이터(2016~2018년 3개년, 3개 공항)로 conv2D 기계학습 수행 결과, 테스트 데이터에 대해 73%의 f1 score를 보인 모델이 가장 좋은 성능을 보였다. 해당 모델로 학습에 사용되지 않은 검증 데이터를 분석한 결과는 [표 4-1]과 같다.

[표 4-1] 민간공항 모델 수행 결과

optimizer = torch.optim.Adagrad(classifier.parameters())						
model = Model_F1(73.36)_Epoch(77)_(created_2021-08-30_210534).pt						
Test 결과	길이	Loss	Accuracy	Recall	Precision	F1 Score
	9,520	0.4948	66.19	91.73	61.13	73.37
Valid 데이터		길이	Accuracy	Recall	Precision	F1 Score
공항/연도	유효시간					
제주19	684	6,480	84.68	78.80	38.86	52.05
제주20	875	6,561	78.17	87.66	36.68	51.72
양양19	423	6,480	82.69	75.89	23.94	36.39
인천19	347	6,480	83.89	78.96	22.01	34.42
여수16	298	6,432	79.09	44.63	10.13	16.51
김포19	149	6,480	83.10	72.48	9.29	16.48
김포20	98	6,561	77.31	84.69	5.33	10.03
울산20	80	6,561	42.08	96.25	1.99	3.89

모델 수행 결과, 학습에 사용된 인천, 제주, 양양공항의 경우, 검증 데이터에서도 급변풍 유효시간 길이에 관계없이 안정적으로 70~80%대의 accuracy와 recall 성능[36]을 보였다. 그리고 유효시간이 적은 데이터일수록 precision이 낮게 나타났고, 따라서 f1 score도 다소 낮게 나타났다.

36) Hon, C.C.(2020)가 개발한 싱가포르 날씨 예측 기계학습 모델에서도 70~80%대의 성능을 보여주었다.

유효시간이 적은 데이터는 [표 3-7]에서 실제 Positive(TP+FN)이 작다는 의미이다. 실제 Positive 중 TP를 재현해내는 recall의 경우, 유효시간의 크기와 관계없이 70~80%를 보여, TP의 크기는 전체 유효시간에 대해 일정 수준을 재현해내고 있다고 볼 수 있다. 따라서 유효시간이 적었던 연도나 공항의 데이터에서는 그 크기에 비례해 작은 양의 TP를 가지게 된다.

precision은 TP를 맞추기 위해 모델이 얼마나 많은 Positive를 예상했는지로 해석할 수 있다. 즉 TP를 예상하기 위해, 얼마나 적은 FP(false-alarm)을 예상했는지가 precision 성능을 결정하게 된다. 그런데 TP가 그 데이터의 유효시간 크기에 비례하므로, 유효시간이 적은 데이터는 작은 TP 값을 가지고, precision 계산식에서 분자에 해당하는 TP[표 3-7]가 작으므로, 약간의 오경보만으로도 작은 TP 값에 비해 분모에 해당하는 TP+FP가 크게 증가하게 된다. 따라서 유효시간이 적은 데이터일수록 precision이 낮게 나타나는 경향을 보이는 것으로 분석된다.

유효시간이 적은 김포공항 검증 결과에서도, recall은 70% 이상의 성능을 보여, 항공안전을 확보하는 관점에서 놓치는 급변풍 발생 사례는 적을 것으로 보인다. TP가 작은 경우에도, 예보관이 본 모델 결과를 기반으로 모델의 과대모의 경향을 고려하여 급변풍 예상에 착안한다면, 급변풍 예상 시간을 조기에 지원하여 안전을 확보하는 데 유용할 것이다.

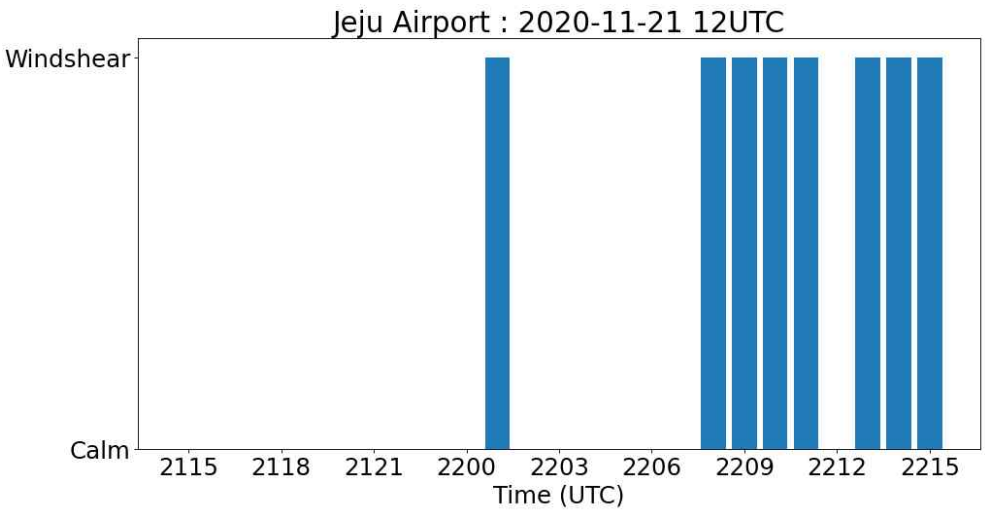
학습에 사용되지 않은 울산공항의 경우에는 유효시간이 80시간으로 극히 낮아 precision과 f1 score 성능이 매우 낮게 나타났으나, recall은 95% 이상의 수준으로 급변풍 발생을 놓칠 확률은 매우 낮을 것으로 보였다. 이에 학습에 사용되지 않은 공군 비행단의 사례에 적용해도, 항공안전 확보 측면에서 급변풍 경보를 조기에 파악할 수 있는 효과를 기대할 수 있을 것이다.

다음은 valid 데이터에 사용한 데이터의 날짜 중 일부 모델 결과를 시계열 자료로 나타낸 것이다.

[표 4-2] 급변풍 경보 발표 기록 일부

공항명	연	월	일시(DDHHMM[37])	
			발효시간(UTC)	해제시간(UTC)
제주공항	2020	11	212209	220100
			221235	221800
			270656	280100

[표 4-2]는 민간공항의 급변풍 경보 발표 기록 중 일부이다. 제주공항에서 2020년 11월 21일과 22일 간 발표되었던 급변풍 기록에 대해 본 연구 모델이 어떻게 예측하는지 [그림 4-1]과 [그림 4-2]를 통해 시계열로 가시화하였다.

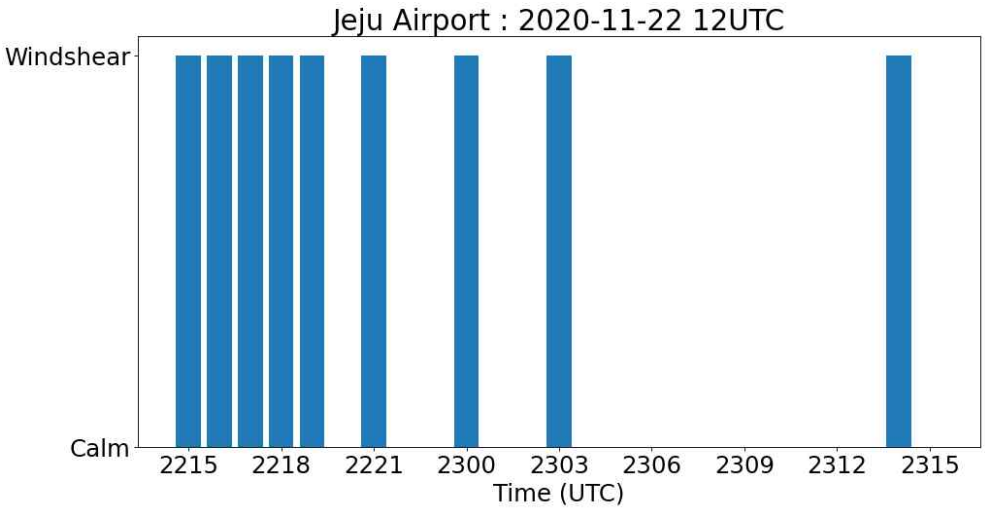


[그림 4-1] 제주공항 2020-11-21 연구 모델 시계열 자료

21일 22시~22일 01시 간 발표된 급변풍에 대해서는 22일 01시 급변풍만 예측하였고, 22일 12~18시 간 발표된 급변풍에 대해서는 모두 잘 예측하였다. 다만 22일 08~11시, 22일 19시 이후에 과대모의가 나타났고, 21일 22시~22일 00시 간에는 과소모의가 나타났다[그림 4-1, 그림 4-2].

37) D : Date(일), H : Hour(시), M : Minute(분)

22일 오후의 경우와 같이 급변풍이 확실히 빈번하게 발생하는 날은 놓치지 않고 잘 모의하면서, 또한 22일 오전의 경우와 같이 2~3시간 정도 한번씩 발생하는 경우에도 과대모의하지만 놓치지 않고 예측해낸다는 점에서 볼 때, 군용기의 안전을 확보하는 관점에서는 활용가능한 수준으로 평가된다.



[그림 4-2] 제주공항 2020-11-22 연구 모델 시계열 자료

[표 4-3] 제주공항 모델 예측 사례[38] 성능 지표

구분		예상	
		Positive	Negative
실제	Positive	7	2
	Negative	9	31

양일간 실제 급변풍이 발생한 시간 9시간 중 모델 recall 성능은 7시간으로 78%, accuracy 성능은 총 49시간 중 38시간으로 78% 수준을 보였다 [표 4-3]. 제주공항 2020년 1년 치 데이터의 검증 결과 accuracy 성능[39]과 유사한 결과를 보였다.

38) 2020년 11월 21일~22일 간
39) [표 4-1] 제주20 Accuracy 성능은 78.17% 였다.

제 2 절 공군 비행단 적용 사례

본 연구에서 개발한 급변풍 예측 모델을 다른 공군 비행단에 적용하여 모델의 성능을 비교/평가하였다. 기존 군 공항에서는 급변풍 경보를 지원하지 않아, 급변풍 경보 대신 강풍특보[40]가 발표되었던 날과 그렇지 않은 날의 군 공항 input 데이터를 생성하여 급변풍 경보 시계열 자료를 생성하였다.

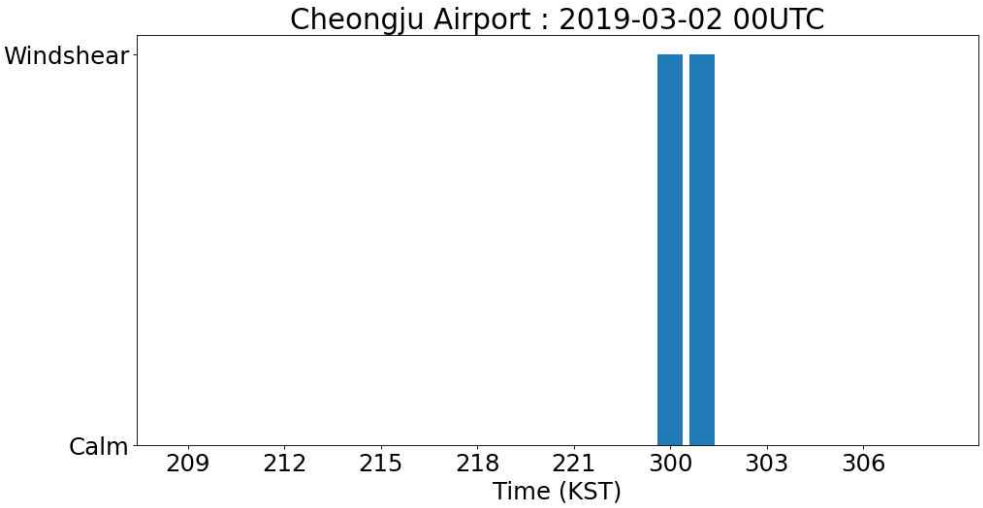
강풍이 불었던 2019년 3월 27일의 경우, 27일 자정 즈음까지 종일 급변풍 경보를 예측하고 있다. 당시 강풍 발생 상황이었으므로, 급변풍의 위험이 있었을 것이다. 또한 precision이 낮아 급변풍을 과대모의 하는지 확인하기 위해, 강풍이 발생하지 않고 상대적으로 맑았던 2019년 3월 2일의 사례와 비교해보았다.

일반적으로 야간 시간대에는 풍속이 약해[41](Zhang, 2015) 모델의 과대모의 가능성을 배제할 수 없지만, 동일 예측시간대의 기상청 LDAPS 시계열 자료와 비교해보면 자정 이후 상층풍은 북서풍으로 바뀌는 반면, 저층 바람은 남동풍이 우세해져 풍향에 대한 급변풍이 강화됐을 가능성이 있는 것으로 예상된다.

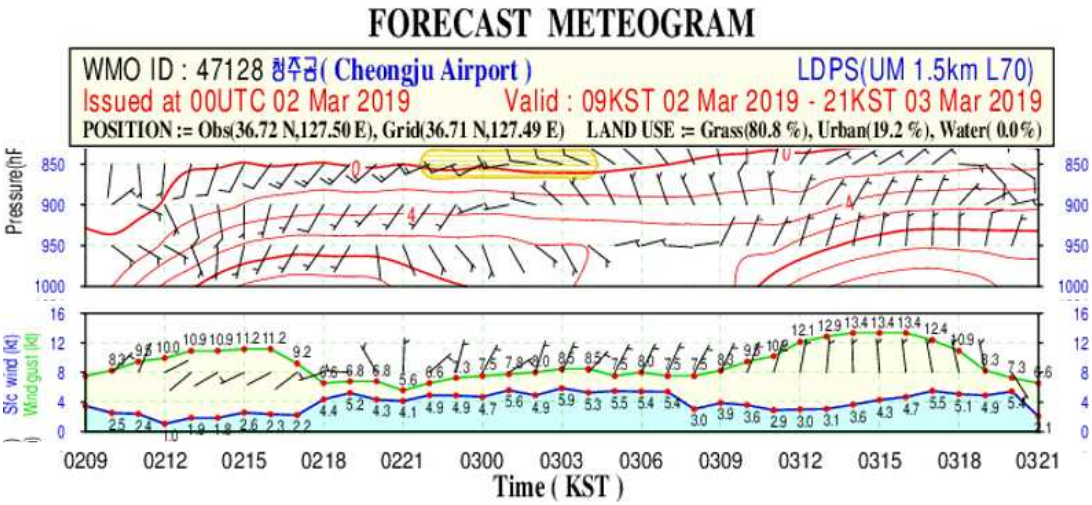
---

40) 강풍주의보 또는 경보. 육상에서 순간풍속 38kts(19m/s) 이상이 예상될 경우, 주의보를 발표하도록 규정하고 있다(공군본부, 2019, 「기상업무」).

41) Chapter 10 Boundary Layer Winds-10.3.3 Diurnal Variations-Figure 10.11



[그림 4-3] 청주공항 2019-03-02 연구 모델 시계열 자료

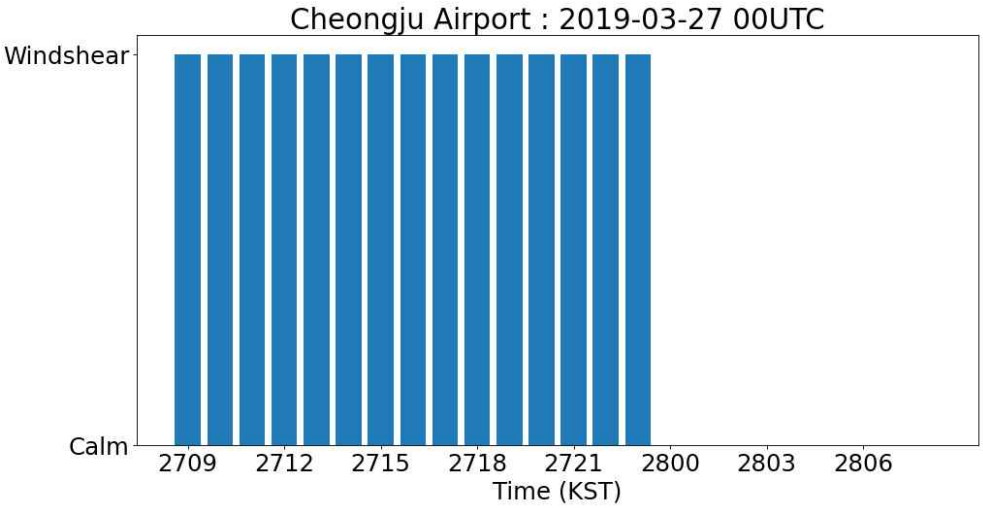


[그림 4-4] 청주공항 2019-03-02 기상청 LDAPS 시계열

2019년 3월 2일 청주공항 사례의 경우, 본 연구의 모델에서는 3일 00~01시 간 급변풍 경보를 예측하고 있다[그림 4-3]. 이 날 전체적인 풍속은 강하지 않았으나, 연구 모델에서 급변풍 경보를 예측한 시간대에 기상청에서 제공한 LDAPS 시계열 자료[그림 4-4]를 확인해보면, 2일 23시 이후 지상풍은 북동풍이 우세해지고, 950hPa 저고도 상층풍은 남동풍 풍속이 증가하며 우세해지는 것을 볼 수 있다. 그리고 3일 03시 이후에는 상층풍이

점차 동풍으로 바뀌며, 지상풍과의 풍향 차이가 감소하는 것을 확인할 수 있다.

풍속이 약한 날이었으므로 본 연구의 모델이 과대모의했을 가능성을 배제할 수 없지만, 모델이 저고도 풍향 차이에 의한 급변풍을 잘 예측하고 있고, 또 대부분의 시간에는 급변풍이 없었음도 잘 모의하고 있어, 모델이 예측한 시간대에 대해서 급변풍이 발생하는지를 예보관이 더 세부적으로 분석하기 위한 결정을 내릴 근거를 마련할 수 있다.



[그림 4-5] 청주공항 2019-03-27[42] 연구 모델 시계열 자료

이 날은 청주 비행단에 강풍특보가 발표된 날이며, [그림 4-5]에서 본 연구 모델은 09~23시까지 종일 급변풍 경보를 예측하고 있다. 강풍이 발생하는 날에는 인접한 고도 간에 작은 풍향 차이로도 큰 급변풍을 유발할 수 있으므로 적절하게 모의하고 있다고 판단된다.

기계학습에 적용하지 않은 공군 비행단 사례에서도, 강풍 발생일에 관해 본 연구 모델이 급변풍을 잘 예측하고 있었다. 강풍이 발생한 날에는 높은 급변풍 발생 가능성을, 유사한 시기에 그렇지 않은 날에는 급변풍을 예측하지 않음으로써 예보관과 사용자가 급변풍 예측 정보에 대한 신뢰를 얻을 수 있었다.

42) 당일 청주공항 순간최대풍속 39kts(20m/s)

강풍이 발생하는 날에는 조금의 풍향 변화로도 큰 변화경향을 얻게되고, 마찰이 심한 지면과 저고도에 가까울수록 변화경향이 커지게 되므로, 급변풍 정보가 전무했던 공군 비행단에서는 강풍 예보와 급변풍 가능성을 함께 토의하곤 하였다. 실제 관측 데이터가 없어 제4장 제1절 모델 수행 결과와 같이 정량적인 성능 비교는 불가능했지만, 본 연구 모델 예측 결과와 조종사 보고를 비교하며 모델 성능을 자체 평가하고, 조종사로부터 피드백을 받아 개선해 나갈 수 있을 것이다.

제 3 절 토 의

본 연구는 LDAPS 분석장과 예상장에서 풍향속, 기압, 연직속도 등 요소와 MODIS 위성자료의 Land cover type 정보를 추출하여 데이터를 구성하고, 급변풍을 예측하는 기계학습을 진행하였다. 연구에 사용된 다양한 요소가 각각 다른 가중치로 학습에 사용되었으며, 학습 과정에서 급변풍에 가장 큰 영향을 미친 요소는 풍향속이었고, 특히 단순 풍속보다는 활주로에 대해 상대적인 풍향을 적용한 풍속 벡터를 사용했을 때 모델의 성능이 향상되었다. 풍향속 외에는 Land cover type 자료를 추가했을 때 모델의 성능이 크게 향상되어, 급변풍 발생에 지형이 중요하게 작용하고 있음을 확인할 수 있었다.

본 연구는 약 24시간 사전에 급변풍에 대한 예보를 생산하고 지원하여, 조기에 위험을 식별하고 항공안전을 보장하기 위함이었는데, 모델의 성능을 향상시키기 위해서는 실시간 관측정보와 급변풍 실황을 연계해볼 수 있을 것이다. 예컨대 AWS 지상풍향속 관측정보는 전국적으로 실시간으로 확인할 수 있으며, 급변풍을 관측하고 있는 공항 지점에서 주변 AWS 풍향속 정보를 바탕으로 바람의 변화에 따른 급변풍 발생 여부를 실시간으로 대조하여 모델의 예측 성능과도 비교해볼 수 있을 것이다. 이는 현재 급변풍 관측을 실시하지 않고 있는 군 공항에 대해서도 주변 AWS 정보는 활용 가능하므로, 민간공항과 동일한 기준으로 모델 성능을 비교해볼 수 있다는 점에서 추후 군 공항에서의 모델 성능 비교/검증 과정에서 필수적으로 검토해야 할 방법 중 하나이다.

지상관측망과 더불어 공항 주변의 기압배치 등 종관분석 군집분류(clustering)를 적용해볼 수도 있을 것이다. 기압배치와 같은 종관장은 지역과 계절, 시간 등에 영향을 받을 수 있기 때문에, 종관장을 반영한 군집분류는 학습에 사용되는 데이터를 효과적으로 분류하고, 각 경우에 따라 효율적인 기계학습으로 모델 성능의 향상을 기대할 수 있을 것이다.

제 5 장 결 론

본 연구를 통해 공군은 전국의 비행단에 통합 적용하여 24시간 이내의 급변풍 경보를 예측할 수 있는 모델을 개발했다. 이는 공군이 저고도 급변풍 정보를 독자적으로 획득할 수 있는 최초의 분류모델이다. 본 연구의 급변풍 예측 모델은 기상예보분야 업무에서 사용되고 있는 수치모델자료를 사용해 급변풍을 예측하고 있고, 성능 결과에서도 각각 78% 수준의 accuracy와 recall 성능을 보여 항공안전 측면에서 활용하기에 충분한 수준이므로, 실무에 적용하여 활용될 가능성을 보였다. 본 연구 모델은 전국 3개 민간 공항에 대한 수치모델자료와 위성자료를 통합하여 기계학습을 시켰기 때문에, 개발된 모형을 지역에 국한되지 않고 공군 비행단에도 일반화할 수 있었다. 따라서 전국의 모든 공군 비행단에 본 연구의 모델을 적용할 수 있을 것으로 판단되며, LDAPS 자료와 더불어 현재 공군이 운영하고 있는 KAF-WRF 자료를 활용하여 급변풍 정보를 즉시 추가할 수 있다.

본 연구 모델은 지역에 국한되지 않는 통합 모델이므로, 위치가 바뀌는 항공모함이나 우리가 접근할 수 없는 북한, 국외 지역에 대한 정보도 생산할 수 있다. 그리고 전투기는 광범위하고 다양한 고도의 공역을 사용하므로, 공군 조종사 보고(PIREP, Pilot Reports)와 연결지어 저고도 뿐 아니라 고고도 난류 정보 생산에도 확대 적용할 수 있다. 기계학습에 사용한 데이터의 종류를 기상 데이터 외의 종류로 확대하면, 급변풍 뿐 아니라 모든 지상 Go-around[43] 상황에도 적용할 수 있을 것이다.

---

43) 착륙복행. 보통 조종사가 자신의 판단에 따라 진행하지만 활주로 또는 항공교통 상황 등의 사유로 도착기 진입의 지속이 안전하지 않다고 판단될 때에 관제사의 지시를 받기도 한다.

참 고 문 헌

고경규, 샤자드, and 정은성(2021), “기상정보 예측 성능 개선을 위한 빅데이터 병합 및 딥러닝 모델 최적화” 「전자공학회논문지」, Vol.58(5), p.39-46

김선애, 김맹기, 박종철, 장동호, Seonae Kim, Maeng-Ki Kim, Jongchul Park, and Dong-Ho Jang(2015), “기후변화 시나리오를 활용한 한국의 미래 상층풍의 바람 특성 분석” 「기후연구」, Vol.10(3), p.273-285

공군본부(2017), 「항공기상관측」, 공본지침서 16-1-1(2017. 1. 1.)

공군본부(2019), 「기상업무」, 공군규정 16-1(2019. 12. 10.)

국토교통부 항공정보서비스(AIS), Instrumental approach Chart, ais.casa.go.kr

기상청(2011), 「태풍백서」

기상청 기상자료개방포털, data.kma.go.kr

“머신러닝 기본 용어 정리”, https://jjeongil.tistory.com/697

방재기상정보시스템, afso.kma.go.kr

“Deep Learning - 학습관련 기술” https://blog.naver.com/lego7407/221681014509

"[Tensorflow] Class Imbalance 문제 대처하기(WCE, Focal Loss)" https://blog.naver.com/siniphia/221764461342

한국공항공사, airport.co.kr

항공기상청(2015), 「항공기상업무지침」, 제9차 개정판

항공기상청(2021), 「공항경보 및 급변풍경보 지침」

항공기상청, amo.kma.go.kr

Branco, P., Torgo, L., Ribeiro, R.P.(2015), "A survey of predictive modelling under imbalanced distributions", *CoRR*

Estabrooks, A., Jo, T., Japkowicz, N.(2004), "A multiple resampling method for learning from imbalanced data sets", *Computational Intelligence*, Vol.20(1): p.18-36

Fu, Donglei, Yuping Liu, Haizhi Li, Songtao Liu, Bo Li, Samit Thapa, Stephen Yabo, Xiazhong Sun, Bo Tang, Jinxiang Zuo, Hong Qi, and Chongguo Tian(2020), "Evaluating the Impacts of Land Cover and Soil Texture Changes on Simulated Surface Wind and Temperature", *Earth and Space Science*, Vol.7(9)

Gidaris, S., Singh, P., Komodakis, N.(2018), "Unsupervised Representation Learning by Predicting Image Rotations", *International Conference on Learning Representations*

Golding(2005), “Low-Level Windshear and Its Impact on Airlines”, *Journal of Aviation/aerospace Education and Research*, Vol.14(2)

Holton(2004), “An Introduction to Dynamic Meteorology”, *International Geophysics Series*, Vol.88(4th eds.)

Hon, Chua Chin(2020), “Weather Predictions: ‘Classic’ Machine Learning Models Vs Keras”, https://towardsdatascience.com/weather-predictions-classic-machine-learning-models-vs-keras-41604f8553e2

Hon, Kai-Kwong(2019), “Predicting Low-Level Wind Shear Using 200-m-Resolution NWP at the Hong Kong International Airport”, *Journal of Applied Meteorology and Climatology*, Vol.59(2): p.193-206

“The History of Convolutional Neural Networks”, https://glassboxmedicine.com/2019/04/13/a-short-history-of-convolutional-neural-networks/

ICAO(2005), *Manual on Low-level Wind Shear*

Iijima, T, Matayoshi, N, and Ueda, S., “Operational Concept and Validation of a New Airport Low-level Wind Information System”, *Aeronautical Journal*, Vol.124(1277), p.975-1015

Jeong, Dae Il, and Sushama, Laxmi(2019), “Projected Changes to Mean and Extreme Surface Wind Speeds for North America Based on Regional Climate Model Simulations”, *Atmosphere*, Vol.10(9), p.497

Johnson, Justin M, and Taghi M Khoshgoftaar(2019), "Survey on Deep Learning with Class Imbalance." *Journal of Big Data* Vol.6(1), p.1-54

Kim, Dong-Ju et al.(2020), “Characteristics of LDAPS-Predicted Surface Wind Speed and Temperature at Automated Weather Stations with Different Sounding Land Cover and Topography in Korea”, *Atmosphere*, Vol.11(11), p.1224

Li, Yupeng, Yaning Chen, and Zhi Li(2019), “Effects of Land Use and Cover Change on Surface Wind Speed in China”, *Journal of Arid Land*, Vol.11(3), p.345-356

Lin, C. et al.(2021), “Overview of Low-Level Wind Shear Characteristics over Chinese Mainland”, *Atmosphere*, Vol.12(5), p.628

Liu, B., Tsoumakas, G.(2019), “Dealing with class imbalance in classifier chains via random undersampling”, *Knowledge Based Systems*, Vol.192, p.105-292

Makhoul, John et al.(1999), “Performance Measures For Information Extraction”, *Proceedings of DARPA Broadcast News Workshop*

NASA, Earthdata, search.earthdata.nasa.gov/search

Sasaki, Yutaka(2007), “The Truth of the F-measure”, *Teach Tutor Mater, School of Computer Science, University of Manchester*

Tomek, I.(1976), “Two modifications of CNN”, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.SMC-6(11): p.769-772

Weiss, G.M., Provost, F.J.(2003), “Learning when training data are costly: The effect of class distribution on tree induction”, *The Journal of Artificial Intelligence Research*, Vol.19, p.315-354

Yann LeCun et al.(1998), “Gradient-Based Learning Applied to Document Recognition”, *IEEE*, Vol.86(11): p.2278-2324

Zhang, Matthew Huaiquan(2015), “Wind Resource Assessment and Micrositing : Science and Engineering”, China Machine Press, Wiley, Singapore

Abstract

Development of Windshear Predicting AI Model for Air Force Support

Joon Young Kwak
Department of Environmental Planning
Environmental Management Major
The Graduate School of Environmental Studies
Seoul National University

Windshear is one of the biggest threats in aviation. Windshear indicates the rapid changes of wind, so it threatens aviation safety significantly, while the information of windshear is provided only to civil aviation currently. As Air Force aircraft must run the flights for the national security, even under hazardous weather conditions, therefore it is urgent to obtain windshear information of military airbase. Even there are no observation devices for windshear in military runways, this study developed windshear predicting AI model based on civil aviation windshear data to secure the information for the military runways as well. To accumulate and to integrate the past data from 3 civil airports(Incheon, Jeju, and Yangyang), this study decided to use convolution-2D machine learning. Input data is

composed of weather data and non-weather data; weather data is composed of low level wind direction/velocity, vertical velocity, and geopotential height from LDAPS numerical model, while non-weather data is composed of land-cover type from MODIS satellite. The input data is randomly split into 4:1 ratio for train/test data respectively. As a result of the study, the model performed 70~80% accuracy by Confusion Matrix. In accordance with this performance, Air Force forecasters would be able to plan the flights avoiding dangerous timings by windshear predicting model results.

keywords : Windshear, Air Force, Artificial Intelligence, Machine Learning, Numerical Model, MODIS

Student Number : 2020-26319