

# 군사과학연구

Journal of Military Science and Technology Studies

ISSN 1975-3888  
제16권 제1호 2023년 6월



## 연구논문

사례를 통해 살펴본 M&S 원시자료변환(Data Wrangling) 방안

윤봉규

기상정보를 활용한 북한 탄도미사일 발사 가능 환경 예측 모델 연구

조성일 · 이춘주 · 문호석

AHP(Analytic Hierarchy Process)를 활용한  
프로젝트 소과제별 중요도 평가 방법

강승욱 · 윤진성 · 소준호 · 이용복

설문조사 시 특이한 답변 결과에 대한 원인 분석 사례 연구  
- 범국민 안보 의식조사 한미연합훈련의 필요성 답변 결과를 중심으로 -

윤영빈 · 문호석



국방대학교  
국가안전보장문제연구소

# 군사과학연구

Journal of Military Science and Technology Studies

ISSN 1975-3888

제16권 제1호 2023년 6월



## 연구논문

- 사례를 통해 살펴본 M&S 원시자료변환(Data Wrangling) 방안 ————— 윤봉규 1
- 기상정보를 활용한 북한 탄도미사일 발사 가능 환경 예측 모델 연구 ————— 조성일 · 이춘주 · 문호석 17
- AHP(Analytic Hierarchy Process)를 활용한 프로젝트 소과제별 중요도 평가 방법 ————— 강승욱 · 윤진성 · 소준호 · 이용복 31
- 설문조사 시 특이한 답변 결과에 대한 원인 분석 사례 연구 ————— 윤영빈 · 문호석 41  
- 범국민 안보 의식조사 한미연합훈련의 필요성 답변 결과를 중심으로 -



국방대학교  
국가안전보장문제연구소



## Research Papers

- Case Study Approach to Data Wrangling Techniques for Modeling and Simulation  
/ **Bongkyoo Yoon** ..... 1
- A Study on Predicting the Readiness Posture Environment of North Korea's Ballistic Missile using Weather Information  
/ **Sungeel Jo · Choonjoo Lee · Hoseok Moon** ..... 17
- Method of evaluating the importance of tasks in a project using AHP(Analytic Hierarchy Process)  
/ **Seunguk Kang · Jinseong Yun · Junho So · Yongbok Lee** ..... 31
- A Case study of analyzing the cause of unusual survey responses  
- Focusing on the necessity of the ROK-U.S. Joint Training in the national security consciousness survey -  
/ **Yeongbin Yun · Hoseok Moon** ..... 41



사례를 통해 살펴본  
M&S 원시자료변환(Data Wrangling) 방안

Case Study Approach to Data Wrangling Techniques for  
Modeling and Simulation

윤봉규<sup>1)</sup>  
Bongkyoo Yoon

ABSTRACT

This paper provides an overview of necessary data wrangling techniques in the defense area, where Modeling and Simulation (M&S) approaches are increasingly popular. The exponential growth in computing power across various domains has had a profound impact on the use of M&S in defense. However, there is a dearth of necessary techniques for dealing with M&S model data in the defense domain. In response to this gap, the present study explores data wrangling techniques leveraging the R tidyverse package, using two case studies from a Markovian model and an ABM simulation model that are relevant to defense-related analyses. The techniques introduced in the study have important implications for improving data manipulation practices for insightful M&S analyses among researchers and practitioners in the defense sector.

Keyword: R Tidyverse, Data Wrangling, Modeling and Simulation(M&S)

## 1. 서론

최근 주목 받고 있는 AI 분야의 발전은 컴퓨팅 파워의 급격한 증가에 의해 가능해졌으며, 날씨 예측, 석유 탐사, 신약 개발 등의 다양한 영역에서 컴퓨팅 파워의 증가는 획기적인 변화를 가져왔다. 이런 분야에서 49%에서 94%의 개선이 컴퓨팅 파워의 증가에 따른 성과라는 주장 있을 정도로 컴퓨팅 파워 즉, 계산 능력의 증가는 사회 전반의 변화를 주도하고 있다(Zewe, 2022).

컴퓨팅 파워의 급격한 증가는 연구방법론에도 변화를 가져왔다. 기존에는 계량적 연구 방법으로 방정식을 세우고 해를 구하는 방법에 기반한 분석적 방법(Analytic Approach)이 시뮬레이션과 같은 계산적 접근법(Computational Approach)에 비해 높은 가치를 가진 것으로 여겨졌으나, 이제는 시뮬레이션으로 대표되는 계산적 방법의 활용 빈도와 중요성이 점점 증가하고 있다. 이 현상은 국방 분야에서도 동일하게 나타나고 있다.

계산적 방법론은 분석 결과가 깔끔한 해(Solution)의 형태로 나오는 것이 아니라 분석적 접근에 비해서 다양한 잡음이 섞인 원시 자료의 형태로 제공될 가능성이 높다. 문제는 ‘자료 분석의 80%는 자료를 정제하고 전처리하는 것(Dasu & Johnson, 2003; Wickham, 2014)’이라는 말이 있을 정도로 자료 분석은 태생적으로 잡음을 제거하고 활용 가능한 형태로 만드는, 시간과 노력이 많이 드는 과정임에도 이해 대한 체계적인 소개나 연구가 많지 않다는 점이다.

컴퓨팅 파워의 증가에 따른 데이터 증가와 빅데이터, 데이터 과학자 등으로 대표되는 데이터 분석의 가치 증가는 원시 자료 전처리에 대한 관심과 필요성을 증가 시켰으며, 그 결과 데이터 전문가뿐만 아니라 현장 실무자나 연구자들이 데이터 전처리에 관심을 가지게 되었다.

이런 변화에도 불구하고 원시자료 전처리에 대한 체계적인 연구는 비교적 최근까지도 거의 없었다(Wickham, 2014).

원시 자료의 전처리에 대한 체계적인 연구가 거의 없었던 이유는 데이터 전처리는 과거에는 주로 데이터베이스 분야 전문가들만 관심을 가지는 주제였고, 해당 분야 전문가가 아닌 연구자나 실무자가 활용하기에는 내용이 너무 방대했기 때문이다. 이에 반해 분석을 위한 자료 전처리는 경험이나 배경 지식에 따라 소요되는 시간의 편차가 크기는 하지만, 시행착오를 거치면서 수행할 수 있기 때문이기도 하다. 결과적으로 특별한 경우가 아니라면 원시 자료의 전처리는 연구자나 실무자에게 시간은 많이 들지만 따로 체계화할만한 가치가 있어 보이지는 않은 분야였다.

또한, 자료 전처리 과정은 사용하는 언어와 컴퓨터에 따라서 다양한 문제를 노출하고, 논리적으로 설명되지 않는 자연어와 유사한 불규칙성이 많아, 표준화나 체계화가 가능한 과학이라기 보다 임기응변으로 상황 상황에 따라 대응하는 기술의 영역으로 치부하는 경우가 많았다. 그 결과 자료 전처리는 숙련된 연구자라고 하더라도 원시 자료의 특성이나 분석 분야에 따라서 전처리 절차나 기법이 차이가 나는 경우가 많다. 이런 이유로 자료 처리 수요가 높아지고 저변이 넓어지는 상황에서도 이에 대한 표준화나 전처리 과정의 효율성을 높이는 방안에 대한 연구가 미흡했다.

최근에는 자료 처리의 관심과 저변 확대에 대한 반향으로 효율적인 자료 전처리와 표준화에 대한 연구가 급속하게 발전하고 있으며, R의 ‘tidyverse’ 패키지는 이런 노력의 일환이다.

‘tidyverse’ 패키지는 자료의 입력을 위한 ‘readr’ 패키지, 문자열처리를 위한 ‘stringr’, 데이터 정제 및 변형을 위한 ‘dplyr’, ‘tidyr’ 등의 패키지를 묶어 놓은 R 통합 패키지이다. 이를 주도한 Wickham은 자료의 시각화에 대한 표준

적인 절차가 존재한다는 Wilkinson(1999)의 주장을 'ggplot2'로 구현할 정도로 시각화(Visualization)를 포함한 자료 처리의 표준화가 가능하다는 생각이 확고한(Wickham, 2010) 데이터 과학자이다.

한편, 'Wrangling'은 원시자료를 시각화(Visualization)와 모형구축(Modeling)에 활용할 수 있는 형태로 정리하는 것을 의미한다(Wickham and Grolemund, 2017). 'Wrangling'은 원시 자료를 처리해서 통찰력 있는 분석 결과를 얻기 위해, 시각화에 바로 활용할 수 있는 형태로 변환하는 것을 포괄하는 개념으로 '원시 자료 전처리와 자료 정제'를 포함하는 개념으로 이해할 수 있다. 본 연구에서는 '원시자료변환'이라는 용어를 전처리와 변환을 포함하는 'Wrangling'의 의미로 사용한다.

본 연구에서는 최근의 R의 'tidyverse' 패키지를 중심으로 진행되고 있는 원시자료 변환 방안(Wrangling)을 기초로, 국방분야에서 M&S를 활용한 분석을 수행할 때 자주 활용되는 원시자료변환 기법을 소개하고자 한다. 특히 문자열을 활용하여 다양한 조건에 부합하는 자료를 추출하는 방식과 최근의 개선된 R 명령어를 활용하여 원시자료변환 과정에서 엑셀로 작업하는 것과 같은 직관적인 방식을 소개함으로써 숙련된 전문가가 아니더라도 복잡한 조건을 만족하는 데이터를 추출하여 원하는 형태로 만드는 방법을 제시하고자 한다.

본 연구는 데이터 전문가가 아닌 국방분야 M&S 연구자나 실무자가 R에 대한 기본적인 이해만 가지고 있더라도 원시자료변환을 용이하게 수행할 수 있도록 사용빈도가 높은 핵심적 원시자료변환 기법을 제시하여, 현장에서 바로 활용할 수 있도록 하는 것을 목표로 한다.

이를 위해 2장에서는 M&S 분석과 관련하여 활용될 사례를 소개하고, 3장에서는 원시자료변환을 위해서 필요한 R의 자료형을 개별 자료와 자료의 집합으로 구분하여 소개하고, 이를 다루

는 법에 대해서 살펴본다. 4장에서는 'tidyverse' 패키지를 중심으로 원시자료변환에 자주 활용되는 기법을 사례를 통해 살펴봄으로써 M&S 자료 전처리에 대한 이해의 깊이를 더하고자 한다.

## 2. 원시자료변환 분석 사례

본 연구는 이해의 편의를 위해서 국방분야 M&S 사례를 통해 도출된 원시자료를 중심으로 원시자료변환 기법을 살펴본다. 본 연구에 활용할 사례는 모델링 분야에서 자주 활용되는 방법론인 마코프체인을 활용한 모델링과, 시뮬레이션 방법인 에이전트기반 시뮬레이션 모델링(Agent Based Modeling)을 통해 얻은 다음과 같은 원시자료를 바탕으로 한다.

### ① 단계형 정비 모형(마코프체인 모델링)

사전검사, 정비, 출고검사의 단계를 거치는 정비 설비의 운영 성과를 비교하기 위해서 마코프체인으로 모델링하는 경우 원시 자료는 보통 아래 <그림 2-1>과 같이 두 테이블로 나타난다.

A data.frame: 396 × 4				A data.frame: 396 × 2	
S1	S2	S3	Index	Index	Prob
<int>	<int>	<int>	<int>	<int>	<dbl>
0	0	0	1	1	0.0008670620
0	0	1	2	2	0.0018991894
0	0	2	3	3	0.0033288712
0	0	3	4	4	0.0001890095
0	0	4	5	5	0.0055546257
0	0	5	6	6	0.0009027354
0	1	0	7	7	0.0003486690
0	1	1	8	8	0.0009073764
0	1	2	9	9	0.0003500549
0	1	3	10	10	0.0003542620

<그림 2-1> 단계형 정비 모형 원시자료(왼쪽: 상태 대응표, 오른쪽: 안정상태확률)

<그림 2-1>은 정비 설비의 사전 검사팀, 정비팀, 출고검사팀이 각각 5, 10, 5팀으로 구성되어 있을 때, 각 정비 단계별 팀별 안배의 적정성을 평가하고자 하는 목적으로 모형을 구축하여 분석한 결과이다. S1은 사전검사 단계에서 현재 업무를 수행중인 팀의 수, S2는 정비업무 수행중인 팀의 수, S3은 출고검사를 수행중인 정비팀의 수를 나타낸다. (S1, S2, S3)가 모여 마코프체인을 구성하며, Index는 (S1, S2, S3)의 일련 번호이다. <그림 2-1>의 오른쪽 테이블은 마코프체인 모형을 분석한 결과로 도출한 각 상태에 머문 시간의 상대적인 비율(안정상태확률)을 나타낸다.

마코프체인을 활용한 모델링 분석 결과가 이렇게 두 개의 표로 주어지는 이유는 마코프체인으로 분석할 때 안정상태확률 계산을 위해 행렬을 활용하기 때문이다. 행렬의 각 행과 열은 (S1, S2, S3)과 일대일로 대응되지만 행렬의 행과 열은 자연수 값을 가지므로 세 개의 숫자 조합으로 이루어진 (S1, S2, S3) 상태를 직접 표현하기 어렵다. 따라서 행렬의 인덱스와 실제 상태의 대응관계를 나타내는 상태 대응표와 분석 결과값인 안정상태확률을 잘 조합해야만 원하는 성과지표를 산출할 수 있다. 예를 들어, 정비팀은 전원 업무중인데 사전검사와 출고검사는 절반 이하로 일을 하는 비율이 어느정도 인지를 알기 위해서는 상태 대응표에서 S2는 10이고 S1과 S3는 0, 1, 2인 (0, 10, 0), (1, 10, 2) ... (2,10,2) 상태의 Index를 확인해서 그 확률 값을 더해야 한다. 즉 두 표를 Index를 기준으로 연결하고 연결된 값을 더하는 과정이 필요하다.

② 미사일 방어시스템 배치  
(에이전트기반 시뮬레이션)

다음 <그림 2-2>는 에이전트기반 시뮬레이션 언어인 NetLogo를 활용하여 왼쪽에 위치한 홍군이 쏜 미사일을 오른쪽에 배치한 청군의

미사일 방어체계가 방어하는 시뮬레이션 결과이다. who는 에이전트(미사일)의 일련번호이며, color는 홍군(15), 청군(105)을 의미한다. xcor, ycor는 에이전트(미사일 또는 미사일 방어체계)의 현재 위치 좌표이며, Time은 시뮬레이션 시간이다. 미사일 방어 시뮬레이션 결과가 각 에이전트(미사일)의 시간대별 위치로 주어진 것이다.

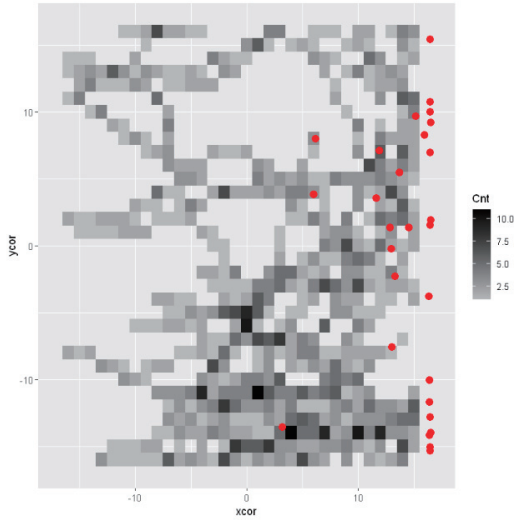
A spec\_tbl\_df: 4972 × 5

who	color	xcor	ycor	Time
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
0	15	-15	-6.1850115	0
1	15	-4	5.3104675	0
2	15	-9	-11.4228603	0
3	15	-2	14.9425713	0
4	15	-2	-6.8643591	0
5	15	-13	5.4798548	0
⋮	⋮	⋮	⋮	⋮
85	15	15.3058922	9.9813362	61
91	15	16.4316808	15.4047195	61
118	105	5.5171209	-9.6121732	61

<그림 2-2> NetLogo 시뮬레이션 결과

다음 <그림 2-3>은 격추되지 않고 청군의 미사일 방어 시스템을 통과해서 피해를 준 홍군 미사일의 이동 경로이다. x축 0을 기준으로 왼쪽이 홍군, 오른쪽이 청군의 미사일 방어시스템이 배치되어 있는 모형이므로 홍군의 공격이 왼쪽에서 오른쪽으로 진행되었다. 빨간색 원은 살아남은 홍군 미사일이 타격한 지점을 나타내며, 경로의 색깔이 진할수록 많은 미사일이 해당 지점(좌표)을 지나갔다는 것을 의미한다. 결과적으로 격추를 피한 홍군의 미사일이 아래쪽을 통과한 경우가 많았다는 것을 알 수 있다.

아래 <그림 2-3>은 앞선 <그림 2-2>의 원시 자료를 격자형으로 그릴 수 있도록 xcor, ycor를 격자의 좌표를 나타낼 수 있도록 실수에서 정수로 위치좌표를 바꾸고, 각 위치좌표를 지나간 미사일의 수를 포함하는 표가 있어야 그릴 수 있다.



<그림 2-3> 격추되지 않은 흉근 미사일의 이동경로와 타격지점

### 3. 자료의 입출력과 R의 자료형

통찰력 있는 분석을 위해서는 앞선 <그림 2-1, 2-2>의 원시자료를 원하는 형태로 변환해야 한다는 것을 앞서 사례를 통해서 확인할 수 있었다. 이제 이를 수행하기 위해서 필요한 배경 지식인 자료의 종류를 나타내는 자료형(Data Type)에 대해서 살펴보자.

자료형을 원시자료처리에 앞서 파악해야 하는 이유를 먼저 살펴보자. 원시자료변환 과정에서 많은 오류의 원인이 잘못된 자료형의 사용임에도 불구하고, 자료형에 대한 설명은 대부분의 문헌에서 간단히 언급되거나 두꺼운 참고 메뉴얼(Reference Manual)과 같이 지나치게 상세히 설명하는 극단적인 방식인 경우가 많다. 그 결과 피상적인 내용만 알거나 필요한

내용만 빠르게 파악하기 어렵다. 이런 이유로 원시자료입력이나 변환시 자료형과 관련된 문제가 생겨도 문제를 파악하기도 어렵고, 파악하더라도 다시 메뉴얼을 들여다봐야 하는 상황이 생긴다. 그 결과 알고 보면 간단한 문제임에도 문제 식별과 해결에 많은 시간이 걸리는 경우가 많다. 여기에서는 이런 점을 해결할 수 있도록 자료형과 관련해서 문제 식별과 해결에 꼭 필요한 내용을 중심으로 R의 자료형에 대해서 살펴본다.

자료형에 대한 설명 이전에 자료를 R에서 분석을 위해서 데이터를 불러오는 자료 입/출력을 간단히 살펴보자.

원시자료는 다양한 형태의 파일로 제공이 되겠지만, 많이 쓰는 형태가 쉼표로 자료의 값을 분리한 CSV파일이다. CSV는 'Comma Splitted Value'의 약자로 말 그대로 쉼표로 분리한 값을 나타낸다. CSV파일은 다음과 같은 명령어로 불러올 수 있다.

```
> read_csv(prob, "result.csv")2)
```

위 명령은 파일명이 result.csv인 CSV파일을 R에 불러와서 prob이라는 변수에 저장하는 명령어이다. 파일을 출력하는 명령어는 write\_csv이며, 쉼표가 아니라 탭이나 다른 문자로 구분된 파일은 read\_delim을 활용하면 불러 올 수 있다.

원시자료를 불러오면 대부분은 잘못된 값이나 결측값이 포함되어 있기 마련이다. 이를 R에서는 NA(Not Available)로 표기하며, 결측값 여부를 판단하는 명령어는 is.na이다. 한편, R의 대부분의 계산 명령어는 na.rm이라는 옵션<sup>3)</sup>을 포함하고 있다. 이 명령어를 활용하면

2) '>'는 R 명령창의 프롬프트이며, 그 뒤에 오는 것이 R의 명령어이다. 한편, R 명령어 실행전에 'tidyverse' 패키지를 다음과 같이 설치하고 활성화해야 여기에서 나오는 명령어를 사용할 수 있다.

```
> install.packages("tidyverse") # tidyverse 설치
> library(tidyverse) # tidyverse 활성화
```

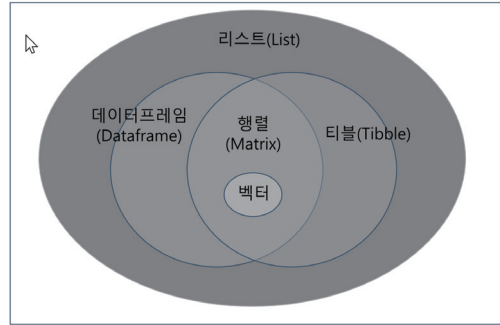
3) 옵션은 명령어 안에서 설정하는 명령어의 상세 내용이라고 할 수 있다. 예를 들어 sum(c(1,NA,3))을 실행

결측값을 제외하고 계산을 수행할 수 있다.

다시 자료형으로 돌아와서, R의 자료형을 살펴보자. R의 자료형은 벡터(Vector), 행렬(Matrix), 요인(Factor), 배열(Array), 문자(Character), 데이터프레임(Dataframe) 등 다양하지만, 기본 자료형은 벡터이다. (Vernables et al., 2023; 노만 매트롭프, 2014).

벡터는 원소의 조합으로 이루어져 있으며, 한 줄로 연결된 상자 안에 숫자나 문자 또는 문장(문자열)이 하나씩 들어가 있는 것으로 비유적으로 생각할 수 있다. 각 상자에 들어가 있는 원소들의 연결이라는 의미로 R에서 벡터는 '이어붙이다'는 의미를 가지는 'concatenate'의 첫자 c를 활용하여 만든다. 예를 들어 c(1,2,3)은 숫자 1, 2, 3이 연결된 벡터이고, c("I am", "You are", "He is")는 세 개의 문자열로 이어진 벡터이다. 벡터는 원소들이 한 줄로 이어진 집합이므로 위치를 나타내는 방식으로 불러올 수 있다. prob[3]은 prob라는 변수의 세 번째 방에 들어가 있는 원소를 불러오라는 명령어이다.

R의 자료형이 혼동을 주는 이유는 원소의 자료형과 원소들의 집합을 나타내는 자료형을 섞어서 사용하기 때문이다. 데이터 처리와 시각화를 위해 만든 언어의 특성상 R은 원소의 자료형보다는 원소들의 집합과 관련된 자료형에 초점을 맞추고 있음에도 기존의 컴퓨터 언어의 자료형을 계승한 측면이 있어 둘이 혼재되어 혼선이 생긴 것이다. 여기에서는 이 둘을 구분하여 원소의 자료형과 집합의 자료형을 구분해서 살펴보자.



<그림 3-1> 집합의 자료형

먼저 R의 원소의 자료형을 살펴보면, 정수(Integer), 실수(Double), 복소수(Complex), 문자(Character)로 나눌 수 있다. 원소의 자료형을 모아둔 집합의 자료형은 원소를 모아둔 포장지나 상자와 같다. 집합의 자료형은 <그림 3-1>과 같이 벡터(Vector)를 기본 단위로 해서 벡터를 여러 개 모은 행렬(Matrix), 행렬에 이름을 붙이고, 데이터처리에 알맞게 만든 데이터프레임(Dataframe)과 데이터프레임과 유사하지만 tidyverse 패키지에 맞게 약간 수정한 티블(Tibble), 마지막으로 여러 자료형을 모아서 묶어 놓기만 한 리스트(List)로 이루어져 있다.

우선 행렬을 살펴보기 위해 벡터를 행렬로 다음과 같이 만들어 보자.

```
> matrix(c(1,2,3,4),ncol=2, byrow=T) (3.1)
     [,1] [,2]
[1,]  1   2
[2,]  3   4
```

식 (3.1)의 명령어 아래쪽은 실행결과이다. 벡터는 특별한 설명이 없으면 아래로 이어져 있는 열벡터를 의미하므로 차례로 불러서 행으로 입력하라는 옵션(byrow=T)을 지정해야 식 (3.1)과 같은 결과를 얻을 수 있다. 식 (3.1) 행렬의 원소는 행과 열의 주소를 통해 불러올 수

하면 결측값 NA를 포함하고 있어 결과 또한 NA가 된다. 대신에 sum(c(1,NA,3), na.rm =T)를 실행하면 결과 4를 얻을 수 있다.

있다. 식 (3.1)의 원소를 변수 ex에 저장했다고 하면 ex[1,2]의 값은 1행 2열의 원소인 2를 불러온다. 원소 호출에 관한 상세한 사항은 Vernables et al.(2023)과 노만 매트롭프(2014)를 참조하기 바란다.

데이터프레임과 티블은 자료 처리에 특화된 형태로 들은 앞선 <그림 2-1, 2-2>와 같은 형식으로 구성되어 있다. <그림 2-1, 2-2>의 자료에서 열은 특성을 나타내고, 행은 개별 관측값을 나타낸다. <그림 2-2>를 기준으로 보면 열은 에이전트의 고유식별번호(who), 색깔(color), x좌표(xcor), y좌표(ycor), 시간(Time)이라는 특성을 나타내고, 행은 이런 특성을 가진 관측값을 나타낸다.

데이터프레임과 티블은 변수명을 정하는 규칙이나, 열의 길이가 다를 때 처리하는 방식, 처음 자료를 만들 때 앞서 지정한 열이름을 그대로 사용할 수 있느냐 여부 등 몇 가지 차이가 있지만(Wickham, 2019), 특별하게 복잡한 자료가 아니라면 처리할 때 큰 차이가 나지 않으므로 여기서는 데이터프레임을 기준으로 설명한다. 들은 거의 완전히 변환 가능하므로 편리한 방식을 선택해서 사용하면 된다.

데이터프레임은 각 열의 이름(변수명)이 있다. 앞선 <그림 2-1>의 왼쪽에 있는 자료는 열의 이름이 S1, S2, S3, Index이고, 오른쪽 자료는 열의 이름이 Index, Prob이다. 열의 이름(또는 변수명)을 불러오는 명령어는 colnames이고 이름을 새롭게 바꾸는 명령어는 renames이다.

아래 식 (3.2a)는 숫자 1,2,3의 원소를 가지는 이름이 x인 자료를 첫 열로, 두 번째 열은 문자 kim, yoon, jung을 원소로 가지는, 이름이 y인 데이터프레임을 ex1이라는 변수에 저장하는 명령어이다. 식 (3.2b)는 데이터프레임 ex1의 이름을 불러오는 명령어이고, 식 (3.2c)는 ex1의 이름을 x, y 대신 Index, Name으로 바꾸라는 명령어이다. 식 (3.2)의 실행 결과는 다음 <그림 3-2>에서 확인할 수 있다.

```
> ex1 <- data.frame(x=1:3, y=c("kim", "yoon", "jung")) ) (3.2a)
```

```
> colnames(ex1) (3.2b)
```

```
> renames(ex1, Index = x, Name = y) (3.2c)
```

A data.frame: 3 × 2		A data.frame: 3 × 2	
x	y	Index	Name
<int>	<chr>	<int>	<chr>
1	kim	1	kim
2	yoon	2	yoon
3	jung	3	jung

<그림 3-2> ex1의 최초 모습(왼쪽)과 변수명(열이름) 변환후(오른쪽)

데이터프레임의 원소는 행렬과 같이 주소로 불러올 수도 있다. ex1[3,2]는 jung이라는 문자열을 불러온다. 한편 ex1[,2]은 ex1의 2열을 불러오는 명령어이며, 이는 행렬의 주소를 활용한 것이다. ex1[,2]은 kim, yoon, jung의 원소를 가지는 벡터를 불러온다. 이는 데이터프레임이 벡터들을 모아놓은 것이라는 점을 명확하게 보여준다. 한편 ex1의 2열의 이름인 Name을 활용해서 ex1\$Name을 통해 불러올 수도 있다. ex1\$Name은 kim, yoon, jung을 원소로 갖는 벡터의 변수명이다. \$는 ‘속에 들어 있는’의 의미로 읽을 수 있다. 즉, ex1\$Name은 ex1 속에 들어 있는 Name이라는 변수를 의미한다.

ex1에 들어 있는 변수 Index와 Name은 ex1이라는 데이터프레임의 포장을 제거할 수도 있는데, attach(ex1)이라는 명령어를 실행하면 데이터프레임 ex1 안에 있는 변수들이 각각의 변수명을 가진 벡터로 풀어 해쳐진다. 이를 다시 묶어서 데이터프레임 안으로 넣으려면 detach(ex1, Index, Name)을 실행하면 된다.

이제 가장 큰 집합의 자료형인 리스트에 대해서 살펴보자. 리스트는 형태와 무관하게 어떤 집합이라도 모을 수 있는 자료형이며 다음 식 (3.3)은 이를 잘 보여준다.

```
> ex2 <- list(x1 = ex1, x2=c(1,2,3), x3=1000)
(3.3)
```

식 (3.3)을 실행하면 ex2에 x1, x2, x3를 원소로 갖는 아래 <그림 3-3>과 같은 리스트가 저장된다.

```
$x1      A data.frame: 3
          x 2
          x    y
<int> <chr>
1     kim
2     yoon
3     jung

$x2      1 · 2 · 3

$x3      1000

<그림 3-3> 리스트 ex2
```

x1에는 ex1에 있던 데이터프레임이, x2에는 숫자 1,2,3으로 이루어진 벡터가, x3에는 숫자 1000이 저장되어 있다. 데이터프레임과 달리 행렬의 형태가 아니며 다양한 형태를 가진 자료를 동시에 포함하고 있다. 여기서 주목할 것은 ex2의 변수명 앞에 \$가 붙어 있다는 점이다. 데이터프레임에서 \$를 ‘속에 들어 있는’이라고 부를 수 있었던 이유는 데이터프레임이 앞선 <그림 3-1>에서와 같이 리스트의 부분집합으로 리스트의 특징을 가지고 있기 때문이었다. ex2\$x1은 ex1과 동일한 데이터프레임이고, ex2\$x2는 1,2,3으로 이루어진 벡터, ex2\$x3은 1000이라는 숫자이다.

리스트는 특이한 성질을 하나 가지고 있는데

원소를 담고 있는 포장지가 두 겹이라는 점이다. 그 결과 원소를 불러오는 방식에 따라서 리스트 변수형을 유지할 수도 있고, 리스트에 저장된 원소의 변수형으로 바뀔 수도 있다. 예를 들어 식 (3.4a)는 데이터프레임을 담고 있는 리스트이지만, 식 (3.4b)는 ex1과 동일한 그냥 데이터프레임이다.

```
> ex2[1] (3.4a)
```

```
> ex2[[1]] (3.4b)
```

식 (3.4b)는 ex2\$x1과 동일하지만 ex2[1]은 담겨 있는 내용은 비슷하지만 리스트 포장지가 아직 남아 있다. 이 차이가 문제를 만들 수 있다. ex[1][,2]는 데이터프레임의 두번째 열에 있는 원소인 kim, yoon, jung으로 구성된 벡터를 불러올 것 같지만, 실행되지 않는다. 반면 ex[[1]][,2]는 kim, yoon, jung으로 구성된 벡터를 불러온다. 포장지가 두 겹인 리스트의 특징으로 인해 한 겹만 풀어서는 그 안에 있는 원소들을 끄집어낼 수 없는 것이다.

원시자료변환 과정에서 의도하지 않아도 사용하는 명령어가 리스트의 형태로 자료를 변환하는 경우가 있다. 이때 리스트와 관련해서 포장지가 두 겹이라는 특징을 이해하지 못한다면 에러를 찾기 어려울 수 있다. 이런 점을 피하기 위해서 리스트를 완전히 벡터로 풀어 일렬로 세우는 unlist 명령이 있다. 데이터프레임의 attach 명령도 일종의 데이터프레임의 포장지를 제거하고 벡터로 만들었던 점을 상기해 보면 데이터프레임이 리스트의 부분집합임을 다시 한번 확인할 수 있다.

이제 자료형들을 변환하는 방법을 알아보자. 먼저 원소의 자료형인 정수, 실수, 복소수는 as.integer, as.double, as.complex 명령어로 변환할 수 있다. 변환할 때 정수, 실수, 복소수의 순서나 역순으로 차례로 변환할 수 있다. 이를 건너뛰어 변환할 수는 없다. 문자형은

as.character 명령어로 변환할 수 있으며 문자를 정수, 실수, 복소수로 변환할 수 있다. 아래 식 (3.5)은 문자 1을 실수로 변환한다. 큰따옴표 안에 있는 “1”은 숫자가 아니며 문자이다. 이 문자는 사칙연산을 하면 에러가 난다. 연산을 위해서는 식 (3.5)와 같이 계산이 가능한 자료형으로 변환해 주어야 한다.

> as.double("1") (3.5)

A data.frame: 3 × 4

x1.x	x1.y	x2	x3
<int>	<chr>	<dbl>	<dbl>
1	kim	1	1000
2	yoon	2	1000
3	jung	3	1000

<그림 3-4> as.data.frame(ex2) 변환 결과

집합의 자료형은 벡터, 행렬, 데이터프레임, 티블, 리스트로 as.vector, as.matrix, as.data.frame, as\_tibble, as.list 명령어로 변환할 수 있다. 그러나 자료형의 변환은 순서가 없지만 자료의 형태에 따라서 변환이 안되는 경우도 있고, 예상하지 못한 결과가 나타날 수도 있으므로 세심한 주의가 필요하다. 예를 들어 as.data.frame(ex2)는 데이터프레임은 행의 수가 일정해야 하므로 차이가 나는 값은 알아서 채우는 형태로 앞선 <그림 3-4>와 같이 변환된다. 이런 변환이 의도한 결과라면 괜찮지만 그렇지 않으면 에러를 만들지는 않더라도 분석자료가 예상과 완전히 달라질 수 있으므로 유의해야 한다.

## 4. Tidyverse를 활용한 전처리 기법

원시자료를 전처리하고 정제하는 원시자료

변환(Data Wrangling)은 자료를 나누고, 나눠진 자료를 다시 변환하고, 합치는 과정을 포함한다(Rattenburry et al., 2017). 이런 맥락에서 여기에서는 Tidyverse 패키지의 명령어를 활용하여 자료를 나누고, 변환하고, 합치는 방법을 살펴본다.

원시자료변환 과정은 변수 이름을 만들거나 이름을 바꾸고 특정 문자를 찾아내는 등의 문자열과 관련된 작업을 반드시 포함한다. 따라서 원시자료변환 과정의 사례를 본격적으로 살펴보기 앞서 공통적으로 필요한 문자열과 변수명을 다루는 법을 먼저 알아보자.

문자열을 다룰 때 가장 중요한 명령어인 paste는 문자를 합치는 명령어이다. paste("a","1")은 "a 1"을 결과로 출력한다. paste는 문자를 합치되 두 문자 사이에 공백을 둔다. 이 공백을 없애기 위해서는 paste0라는 명령어를 활용하면 된다. 이를 응용해서 var1, var1, ..., var9, var10로 명명되는 10개의 변수명을 paste0("var",1:10)을 활용하면 만들 수 있다.

var1에 1,2,...,10을 저장하는 명령어는 var1 = 1:10 이다. 그런데 이 방식이 아니라 다음 명령어를 활용해도 동일한 작업을 수행할 수 있다.

> assign("var1", 1:10) (4.1)

assign은 문자열을 변수 이름으로 해서 자료를 저장하는 명령어다. 식 (4.1)을 수행하면 R에서는 1,2,...,10이 저장된 var1이라는 변수가 생긴다. 여기에서 문자열 "var1"과 변수 var1이 같지 않다는 점을 주목할 필요가 있다. 문자열 "var1"과 변수명 var1의 차이는 문자열 "1"과 숫자 1의 차이와 유사하다. assign은 문자열을 변수 이름을 나타내는 자료형으로 변환하고, 그 안에 자료를 할당(저장)하라는 명령어이다. 문자열을 변수이름으로 바꾸기 위해서는 str2expression이라는 명령어가 필요하다. str2expression은 as.integer나 as.double과 유사

한 변환 명령어이다. 한편, 변수 이름에 들어있는 값을 불러오는 것은 eval이라는 명령어가 필요하다.<sup>4)</sup> 프롬프트 창에 var1이라는 변수 이름을 실행하면 변수에 저장된 자료나 값이 호출되는데, 이 과정은 프롬프트에서 eval을 보이지 않게 수행한다고 생각해도 무방하다. 그러나 문자열에서 변환한 변수 이름은 eval이라는 명령어를 명시적으로 수행해야 저장된 값을 불러온다. 식 (4.2)는 문자열 "var1"에 저장되어 있는 자료를 불러오는 명령어이다. 즉, var1을 프롬프트에서 실행하는 것과 동일하다.

```
> eval(str2expression("var1")) (4.2)
```

식 (4.1)과 (4.2)는 이해하기 쉽지 않고, 심지어 식 (4.2)는 비효율적으로 보이기도 하지만 여러 변수에 여러 자료를 할당하거나 불러올 때 반복문(for loop)를 통해 작업을 자동화할 수 있는 중요한 명령어이다. 이를 확인하기 위해 var1에 10에서 20, var2에는 20에서 30, ..., var10에는 100에서 110을 저장하는 명령을 만들어 보면 다음 식 (4.3)과 같이 같다.

```
> val_list = paste0("var",1:10)
> for (i in 1:10){
  assign(var_list[i],(10*i):(10*i+10))
} (4.3)
```

이 자료를 var\_new1에서 var\_new10에 복사하는 방법은 다음 식 (4.4)와 같다.

```
> val_new_list = paste0("var_new",1:10)
> for(i in 1:10){
```

```
  assign(var_new_list[i],eval(str2expression
    (var_list[i])))
} (4.4)
```

텍스트를 처리할 때 paste와 더불어 사용빈도가 높은 명령어는 특정 조건을 만족하는 문자를 찾아내는 str\_detect이다. 다음 명령식 (4.5)는 "who", "color", "xcor", "ycor", "Time"으로 이루어진 앞선 <그림 2-2>의 NetLogo 시뮬레이션 결과에서 cor이라는 문자열이 들어간 변수 이름을 찾는 명령어이다.

```
> str_detect(c("who", "color", "xcor", "ycor",
  "Time"), "cor") (4.5)
```

str\_detect의 두번째 인수인 "cor"은 정규표현식(Regular Expression)으로 문자열에서 특정 패턴을 찾는 방식으로 널리 활용되는 방식이다. 상세한 내용은 Wickham(2021)과 DataCamp(2022)를 참조하기 바라며, 원시자료 변환에 활용도가 높은 정규 표현식은 <표 4-1>과 같다.

<표 4-1> 간단한 정규표현식

정규표현식	찾는 패턴
^a	첫문자가 a
a\$	끝문자가 a
.a.	중간에 a를 포함한 세 개의 문자
^ak.*b\$	처음에는 ak로 시작하고 마지막이 b로 끝남

식 (4.5)에서 두번째 인수 "cor" 대신에 "cor\$"을 넣으면 끝이 cor로 끝나는 xcor, ycor

4) expression은 R 명령어 표현을 일컫는 자료형이다. 'a <- c(1,2,3)'은 프롬프트에서 입력되면 1,2,3을 a에 저장하라는 R의 명령어이지만, 문자열로 볼 수도 있다. 사람은 상황을 고려하여 이를 쉽게 구분하지만, 컴퓨터는 이를 명확하게 정해주어야 한다. 문자 '1'과 숫자 1을 구분하는 것과 같이 문자열은 문자(텍스트)에 불과하고, expression은 프롬프트에 입력되는 R의 명령을 나타낸다. 이런 맥락에서 eval 명령어는 R 명령어를 입력하고 엔터를 눌러 실행하는 것과 동일한 명령어이다.

을 찾는다. 실제로는 False, False, True, True, False라는 결과를 제시한다.

이제 Tidyverse를 활용해서 본격적으로 원시자료를 변환하는 과정을 살펴보자. 앞서 설명한 바와 같이 원시자료 변환은 자료를 쪼개고, 변환하고, 다시 붙이는 과정으로 이루어져 있다. 이를 위해 우선 자료를 쪼개고 붙이는 과정을 살펴보고, 변환하는 과정은 R4.2.3 버전 이후 새롭게 추가된 `reframe`이라는 명령어를 중심으로 살펴본다.

Tidyverse는 원시자료변환 과정에서 파이프 연산(Pipe Operation)을 사용한다. 파이프 연산은 원하는 데이터를 얻을 때까지 원시자료변환 과정을 직관적으로 이해하기 쉽게 한 단계 한 단계씩 진행하며 앞 단계의 결과를 그대로 활용하는 방식으로 진행되는 자료 변환 기법이다.

이런 과정은 물을 원하는 압력으로 가정에 공급하기 위해서 큰 수도관에서 점차 작고 가는 수도관으로 바꾸면서 압력을 낮추는 과정과 유사하다. 이런 맥락에서 데이터 변환을 차례로 연속적으로 수행하는 방법을 수도(?) 파이프(Pipe) 연산이라고 한다(윤봉규, 2021).

R에서 파이프 연산은 `%>%`를 통해 수행한다. 앞선 <그림 3-2>의 `ex1`은 열이름이 Index와 Name으로 구성된 데이터프레임이다. `ex1`에서 Index가 2 이상인 경우만 뽑는 명령어는 다음과 같다.

```
> ex1 %>% filter(Index >= 2) (4.6)
```

식 (4.6) 뒤에 `'%>% filter(Name == "yoon")'`을 추가하면 `ex1`에서 Index가 2 이상인 자료를 뽑고, 이 자료를 바탕으로 Name이 `yoon`인 자료를 뽑으라는 명령이 된다. 이렇듯 파이프 연산은 순차적으로 자료를 변화시켜 가는 과정을 코드에 반영할 수 있다.

Tidyverse에서 자료를 분리하는 방법은 행을 기준으로 분리하는 방법과 열을 기준으로

분리하는 방법이 있다. 행을 기준으로 분리하는 방법은 식 (4.6)과 같이 `filter` 명령어를 사용한다. 열을 기준으로 분리하기 위해서는 R의 주소체계를 활용한다. 즉 `ex1`의 첫 열은 `ex1[1]`, 두번째 열을 `ex1[2]`를 활용하면 호출할 수 있다. 이를 파이프 연산으로 활용하면 다음과 같다.

```
> ex1 %>% .[1] (4.7a)
```

```
> ex1 %>% .[2] (4.7b)
```

식 (4.7)의 [1], [2] 앞에 붙어 있는 `'.'`은 앞의 변수명 `ex1`을 부득이하게 사용해야 하는 경우 쓰는 약어이다. R의 주소를 부르는 방식은 Tidyverse가 아니라 기본 패키지(Base)에 포함된 명령어이므로 파이프 연산을 완벽하게 구현하고 있지 못하다. 그렇다고 기본 패키지를 바꾸기에는 여러 문제가 발생할 가능성이 있어 타협한 결과로 나타난 것이 이런 표현이다. 주소 체계와 문자열 정규표현식을 응용하면 다음과 같이 `ex1`에서 N으로 시작하는 열만 뽑아낼 수 있다.

```
> ex1 %>% .[str_detect(colnames(),"^N")] (4.8)
```

`str_detect(colnames(),"^N")`는 열이름(`colnames`) 중 N으로 시작하는 열을 찾으라는 명령어이다. `colnames`도 기본패키지에 포함되어 있어 `'.'`를 사용하고 있음을 주목할 필요가 있다. R은 오랜 기간에 걸쳐 여러 사람과 단체에 의해서 개발되어 온 역사성으로 인해서 이런 식의 타협의 결과가 자주 나타난다.

이제 자료를 합치는 방법을 앞선 <그림 2-1>의 단계형 정비 모형 원시자료 사례를 통해 살펴보자. <그림 2-1>의 왼쪽에 있는 상태 참조표를 먼저 만들어 `dat1_1`에 저장하는 명령어는 다음 식 (4.9)와 같다.

```
> dat1_1 <- data.frame(S1 = 0:5) %>%
cross_join(data.frame(S2 = 0:10)) %>%
cross_join(data.frame(S3 = 0:5)) %>%
cbind(data.frame(Index=1:(6*11*6))) (4.9)
```

cross\_join은 원소의 조합으로 순서쌍을 만드는 명령어로 S1의 원소 0,1,...,5와 S2의 원소 0,1,...,10으로 이차원 순서쌍을 만들고 다시, S3의 원소들을 포함한 순서쌍을 만들고 이를 일련번호를 나타내는 Index 변수와 열로 붙이는 (cbind) 코드이다.

식 (4.9)의 코드는 파이프연산을 통해 자료 변환을 순차적으로 진행하면서 변환 과정의 가시성을 높인다는 점을 주목할 필요가 있다. 파이프 연산은 자료의 변환을 블랙박스처럼 결과만 얻는 방법이 아니라 그 과정을 순차적으로 진행함으로써 가시성을 높여 변환 과정의 오류를 줄이는 이점을 제공한다. 한편 두 자료를 옆으로 붙이는 것은 cbind (Column Bind), 아래 위로 붙이는 것은 rbind (Row Bind)를 사용한다.

dat1\_2에 <그림 2-1>의 오른쪽 표인 안정상태확률이 저장되어 있다면, 다음 <그림 4-1>과 같이 두 자료를 Index를 기준으로 합치는 방법은 두 가지가 있다. 두 표의 Index가 순서대로 정렬되어 있는 경우에는 식 (4.10a)를 실행하면 되고, 정렬되어 있지 않은 일반적인 경우에는 식 (4.10b)를 실행하면 된다.

```
> dat1_3 <- dat1_1 %>% cbind(dat1_2[2])
(4.10a)
```

```
> dat1_3 <- dat1_1 %>% full_join(dat1_2,
by="Index")
(4.10b)
```

A data.frame: 396 × 5

S1	S2	S3	Index	Prob
<int>	<int>	<int>	<dbl>	<dbl>
0	0	0	1	0.0008670620
0	0	1	2	0.0018991894
0	0	2	3	0.0033288712
0	0	3	4	0.0001890095
0	0	4	5	0.0055546257
0	0	5	6	0.0009027354
0	1	0	7	0.0003486690
0	1	1	8	0.0009073764
0	1	2	9	0.0003500549
0	1	3	10	0.0003542620

<그림 4-1> 단계형 정비 모형 최종 결과(dat1\_3)

이제 2장의 사례 설명에서 예시로 들었던 상황인 정비팀은 전원 업무중인데 사전검사와 출고검사는 절반 이하로 일하는 비율 계산해 보자. 이를 구하기 위해서는 <그림 4-1>의 자료가 저장되어 있는 dat1\_2에서 S2는 10이고 S1과 S3는 0, 1, 2인 (0, 10, 0), (0, 10, 1) ... (2, 10, 2) 상태를 뽑아서(filter), Prob 변수명을 가진 열을 선택한 후(select) 그 확률값을 다음과 같이 더하면(sum) 된다.

```
> dat1_3 %>% filter(S1 <= 2 & S2 == 10 &
S3 <= 2) %>% select(Prob) %>% sum
```

이제 앞선 <그림 2-2>의 미사일 방어와 관련된 NetLogo 시뮬레이션 결과 자료를 앞선 <그림 2-3>과 같이 홍군의 미사일이 격추되지 않고 지나간 경로를 그리기 위한 자료로 변환해 보자.

A tibble: 544 × 3

xcor	ycor	Cnt
<int>	<int>	<int>
-15	11	4
-13	-16	3
-10	16	1
-3	-13	2
⋮	⋮	⋮
15	0	2
8	-1	1
1	-8	1
10	7	3
11	-9	3

<그림 4-2> 격추되지 않은 미사일의 경로(dat2\_4)

이를 위해서는 1) 격추되지 않은 홍군의 미사일을 파악해서, 2) 그 미사일의 경로 자료를 뽑아서, 3) 미사일이 지나간 경로의 자료를 격자에 적합하게 정수 좌표로 변환하고, 4) 각 격자 좌표를 지나간 미사일의 수를 세어 이를 포함하는 앞선 <그림 4-2>와 같은 표를 만들어야 한다. <그림 4-2>의 Cnt는 각 좌표를 지나간 미사일의 수를 나타낸다. 앞선 <그림 2-2>와 달리 xcor, ycor의 자료형이 실수형을 나타내는 <dbl>에서 정수형을 나타내는 <int>로 변했음을 알 수 있다.

원시자료가 dat2\_15)에 저장되어 있다고 가정하고, 앞선 <그림 4-2>와 같은 표를 만들기 위해 먼저 격추되지 않은 미사일을 파악해 보자. 격추되지 않은 미사일은 청군의 미사일 방어시스템의 작동시간인 60분 이후에도 살아남아 있는 홍군의 미사일이라고 가정하면, NetLogo 시뮬레이션의 원시자료에서 Time이

60보다 크고 홍군의 색깔 코드인 color가 15인 자료를 뽑아내면 된다. 이는 아래 식 (4.10)과 같이 수행할 수 있다.

```
> missed <- dat2_1 %>% filter(color == 15 & Time > 60) (4.10)
```

이제 살아남은 홍군의 미사일은 데이터프레임 missed의 who라는 변수 즉, missed\$who에 저장되어 있다. 이를 활용하여 살아남은 미사일의 경로만 추출하는 방법은 다음과 같다.

```
> dat2_2 <- dat2_1 %>% filter(who %in% missed$who) (4.11)
```

식 (4.11)의 filter(who %in% missed\$who)는 data2\_1의 who 변수가 missed\$who 안에 (%in%) 포함되어 있느냐 여부를 판단해서 뽑아내는 명령어이다.

좌표별로 미사일이 지나간 횟수를 파악하기 위해서는 좌표를 격자형에 맞게 실수에서 정수로 변환해 주어야 한다. 이는 as.integer(round(xcor)) 명령을 사용해서 다음 식 (4.12)와 같이 수행할 수 있다.

```
> dat2_3 <- data2_2 %>% reframe(who, xcor = as.integer(round(xcor)), ycor = as.integer(round(ycor)), Time) (4.12)
```

식 (4.12)의 reframe 명령은 Tidyverse에서 자료를 변환하는 새로운 기법이다. Tidyverse에서 자료를 변환하는 방식은 자료를 추출하고(filter), 변형하고(mutate), 그룹핑하고(group\_by), 변수를 선택해서(select), 요약(summarize)하는 과정으로 이루어진다(윤봉규,

5) 원시자료는 저자에게 요청시 csv파일로 제공되며 >dat2\_1 <- read\_csv("ex2.csv")를 실행하면 dat2\_1에 원시자료가 저장된다.

2021). `reframe`은 `mutate`, `group_by`, `select`, `summarize`를 한 번에 수행할 수 있는 방법이다. 식 (4.12)를 기존 방식으로 수행하기 위해서는 `reframe` 대신 `mutate` 명령어를 사용하면 동일한 결과를 얻을 수 있다.

마지막으로 각 격자 좌표별로 지나간 미사일의 수를 세는 방법은 다음식 (4.13)과 같다.

```
> dat2_4 <- dat2_3 %>% reframe(xcor, ycor,
Cnt = n(), .by=c(xcor,ycor)) %>% unique
(4.13)
```

위 식 (4.13)의 '`Cnt = n(), .by=c(xcor,ycor)`'은 `xcor`, `ycor`를 기준으로 그룹을 만들어 그 그룹의 빈도를 세어서 `Cnt`에 저장하라는 옵션이다. 빈도가 아니라 그룹의 평균을 구하고 싶으면, `n()` 대신 `mean()`을, 최대값은 `max()` 등 그룹의 다양한 특성을 구할 수 있다. 한편, `reframe`의 옵션 `by` 앞에 마침표 점이 포함되어 있음을 주목해야 한다. 이는 앞서 설명한 기술적 타협의 문제이니 여기에서 상세히 설명하지는 않지만 마침표를 포함하지 않으면 명령이 실행되지 않는다는 점은 주목할 필요가 있다. 마지막에 있는 `unique` 명령어는 자료가 중복되는 것을 방지하기 위한 명령어이다. `xcor`, `ycor`는 같아도 `who`가 다르면 다른 자료로 인식되었는데, `who`를 제외하면 동일한 자료의 중복이 된다. 이를 보완하기 위한 명령어로 그룹을 만들어서 자료를 변환한 경우에는 `unique`를 써주는 것이 좋다. 식 (4.13)은 `reframe` 대신 `group_by`와 `summarize` 명령어를 활용해서 다음과 같이 수행할 수도 있다.

```
> dat2_4 <- dat2_3 %>% group_by(xcor,ycor)
%>% summarize(Cnt = n()) %>% unique
```

식 (4.11)-(4.13)는 `NetLogo` 시뮬레이션의 원시자료인 `dat2_1`를 출발점으로 해서, 미사일

방어시스템에서 처리하지 못하고 놓친 홍군의 미사일 자료를 뽑아 `dat2_2`에 저장하고, 미사일이 지나간 경로의 좌표를 정수 좌표로 변환하여 `dat2_3`에 저장하고, 각 좌표 격자에 지나간 미사일의 수를 세어 `dat2_4`에 저장한 세 절차를 수행하는 명령이다. 그런데 각 연산을 차례로 수행하는 파이프 연산의 장점을 활용하여 이를 다음과 같이 한 번에 수행할 수도 있다. 이렇게 함으로써 매 연산을 따로 저장하지 않으므로 연산을 효율적으로 수행하면서도, 세 절차가 어떻게 수행되고 있다는 것을 명시적으로 파악할 수 있는 코드를 작성할 수 있다.

```
> dat2_4 <- dat2_1 %>% filter(who %in%
missed$who) %>%
reframe(who, xcor = as.integer(round(xcor)),
ycor = as.integer(round(ycor)), Time) %>%
reframe(xcor, ycor, Cnt = n(), .by=c(xcor,ycor))
%>% unique
(4.14)
```

한편, `dat2_4`에 저장된 데이터를 활용하여 <그림 2-2>를 그리는 명령어는 다음과 같다. `ggplot`을 활용해서 R 그래프를 그리는 방법은 윤봉규(2021)에 상세하게 소개되어 있으니 참고하기 바란다.

```
> ggplot() +
geom_tile(data=dat2_3,aes(x=xcor,y=ycor,fill=Cnt))
+ scale_fill_gradient(low = "grey", high =
"black")+ geom_point(data=missed,aes(xcor,ycor),
color="red",size=3) + xlim(-17,17) (4.15)
```

## 5. 결론 및 시사점

지금까지 모델링과 시뮬레이션 분야에서 활용도가 높은 방법론인 마코프체인과 에이전트 기반 시뮬레이션 모형에서 자주 나오는 원시자

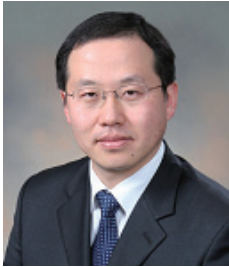
료를 하나씩 사례로 뽑아, R로 원시자료변환(Wrangling)을 수행하는 기법을 살펴봤다. 그 과정에서 원시자료변환에 필요한 배경지식인 자료형과 자료형의 변환, 정규표현식을 포함한 문자형 자료를 처리하는 방법과 이를 R에서 활용하는 방법 또한 살펴봤다.

본 연구를 통해 국방분야 M&S 연구자와 실무자들이 많은 노력을 들이지 않고 원시자료변환을 위한 기본적 기법을 이해하고, 원시자료변환의 의의를 파악함으로써 향후 이 분야의 표준화를 통한 효율성 제고를 통해 국방분야 M&S 발전에 기여할 수 있기를 기대한다.

## 참 고 문 헌

- [1] 윤봉규, "간소화된 그래픽 문법(Grammer of Graphics) 기반 데이터 시각화," 군사과학연구지 14(1): 29-41, 2021.
- [2] 매트로프, 노만 (권정민 역), 빅데이터 분석 도구 R 프로그래밍, 에이콘, 2014.
- [3] DataCamp, "A Guide to R Regular Expressions With Examples," DataCamp, 2022. available: <https://www.datacamp.com/tutorial/regex-r-regular-expressions-guide>.
- [4] Dasu T, Johnson T, Exploratory Data Mining and Data Cleaning, John Wiley & Sons, 2003.
- [5] Rattenbury, T., J.M. Hellerstein, J. Heer, S. Kandel, and C Carreras, Principles of Data Wrangling: Practical Techniques for Data Preparation, O'Reilly, 2017.
- [6] Venables, W.N., D. M. Smith and the R Core Team, Introduction to R:version 4.2.3, 2023.
- [7] Wickham, H., "stringr: Simple, Consistent Wrappers for Common String Operations: Vignettes: Regular Expressions (R package version 1.5.0), 2021. available: <https://cran.r-project.org/web/packages/stringr/vignettes/regular-expressions.html>.
- [8] Wickham, H., Advanced R, Taylor & Francis, 2019.
- [9] Wickham, H. and G. Grolemund, R for Data Science, O'REILLY, 2017.
- [10] Wickham, H., "Tidy Data," Journal of Statistical Software, 50(10), 2014.
- [11] Wickham, H. "A Layered Grammar of Graphics." Journal of Computational and Graphical Statistics 19(1): 3-28, 2010.
- [12] Wilkinson, L., The Grammar of Graphics, Springer, 1999.
- [13] Zewe, A., "Q&A: Neil Thompson on computing power and innovation," MIT News June 24, 2022. available: <https://news.mit.edu/2022/neil-thompson-computing-power-innovation-0624>.

## 저 자 소 개



**윤봉규 (E-mail: [bkyoon1@gmail.com](mailto:bkyoon1@gmail.com))**

1996 연세대학교 경영학사

1998 한국과학기술원 산업공학 석사

2002 한국과학기술원 산업공학 박사

현재 국방대학교 운영분석전공 교수

관심분야 : Agent Based Modeling, Stochastic  
Models in Military O.R, Biz.  
Performance Optimization &  
Innovation.

## 기상정보를 활용한 북한 탄도미사일 발사 가능 환경 예측 모델 연구

### A Study on Predicting the Readiness Posture Environment of North Korea's Ballistic Missile using Weather Information

조성일<sup>1)</sup> · 이춘주<sup>2)</sup> · 문호석<sup>3)</sup>

Sungeel Jo · Choonjoo Lee · Hoseok Moon

#### ABSTRACT

Existing research on North Korea's ballistic missile launches has focused on follow-up management of pattern analysis linking ballistic missiles and nuclear development, as well as tracking and predicting the range of ballistic missile orbits. Most of the predictions related to North Korea's ballistic missile launch were viewed from a political and security perspective related to North Korea. These existing studies can also be seen as dealing with the intentions of the North Korean command in light of various situations. However, even if North Korea intends to launch a ballistic missile, it must also have a weather environment to launch it.

Therefore, this study proposes a model to predict the weather environment of North Korean ballistic missiles using past North Korean weather information and ballistic missile launch information. As a result of the study, random forest and adaboost models among classification models show high prediction accuracy of over 94%. Using the results from a decision tree model to interpret key weather variables that significantly influence ballistic missile launch predictions, it was shown that conditions are unfavorable for missile launch when the wind speed is 2.5m/s or greater. This is in line with the fact that too strong ground winds can negatively impact rocket launches. Such a model, trained on past weather data, can be used to predict suitable weather conditions for ballistic missile launches by inputting new weather data.

Key Words : North Korea's Ballistic Missile, Readiness Posture Environment Prediction models, Weather Information, Machine Learning, Cross Validation, National security

논문접수일 : 2023년 4월 28일, 심사일 : 2023년 5월 31일, 게재확정일 : 2023년 6월 7일

1) 국방대학교 국방과학학과 석사과정

2) 국방대학교 무기체계전공 교수

3) 국방대학교 국방과학학과 교수 / 교신저자(corresponding author)

## 1. 서론

북한의 탄도미사일 발사 여부를 결정하는 것에서 제일 중요한 점은 ‘북한의 의도’이다. 다시 말해서 북한 최고 지도자의 의도에 따라서 발사 여부가 결정될 것이다. 그러나 분명한 ‘발사 의도’가 있더라도 탄도미사일 발사 환경이 조성되어있지 않으면 실행할 수 없다.

본 연구는 북한의 탄도미사일 발사 여부에 관한 예측모델링에 관한 연구이다. 본 연구에서의 예측은 ‘북한 지휘부의 의도’를 예측하는 것은 아니며, 북한이 탄도미사일을 발사할 수 있는 가능한 환경이 조성되었는지를 북한의 기상정보를 통해서 예측하려고 한다. 기상적인 측면에서 발사 환경이 안 되어 있으면 아무리 의도가 있어도 미사일 발사는 불가능하기 때문이다. 본 연구는 북한의 기상정보를 이용해서 북한의 탄도미사일 발사 환경을 예측하려는 측면에 초점을 둔 첫 연구로서 의의가 있다.

북한 탄도미사일의 역사는 1968년 구소련이 제공한 Frog 로켓으로부터 출발한다. 북한은 한미의 우세한 공군 전력을 극복하고 전략적 우위를 확보하고자 탄도미사일 개발에 착수하였다. 1980년대 스커드-B(사거리 300 ~ 500km) 미사일을 개발했고 1990년대 들어선 스커드-C로 사거리를 500km 연장했다[1]. 또한 1990년대 초부터 2,000~2,500km 및 4,000~6,000km의 장거리 탄도미사일인 대포동미사일 1, 2를 개발하기 시작하였다[2]. 2012년부터 신형 액체·고체 추진 탄도미사일에 대한 연구개발을 추진 하여 2016년 개발에 성공한 화성-12형으로 3회에 걸쳐 일본 상공을 통과하는 시험 발사를 하는 등 향후에도 핵 무력 고도화를 위한 미사일 개발과 시험 발사를 지속할 것으로 예상된다[3].

과거 시기별(1984 ~ 2017) 북한 탄도미사일·핵 관련 도발 횟수를 분석해보면 대부분 매년

5월 주변으로 많이 일어난 것을 알 수 있으며, 김일성 시기 8회 중 4회가 5월에 발생했고 김정일 시기 역시 28건 중 5월이 8건으로 가장 많았다. 김정은 집권 이후 또한 2017년까지 전체 76회의 도발 중 42회(55%)가 3 ~ 5월 사이에 일어나며 유사한 결과를 보여주었는데[4] 이를 통해 북한이 계절적 요소를 고려하여 핵 및 미사일 개발을 추진한다고 유추할 수 있다. 즉 ‘북한의 지리적 특성과 겨울철에 매우 낮은 기온 등의 환경 요소를 고려한 것’이라고 볼 수 있다.

하지만, 2022년도에 북한은 이전과는 완전히 다른 탄도미사일 패턴을 보여주었다. <표 1-1>은 '22년 월별 북한의 탄도미사일 도발 현황을 보여주는데 10월과 11월에 각각 6회로 윤석열 정부 출범 후에는 우리나라와 미군의 훈련간 많은 탄도미사일을 발사했음을 알 수 있다.

<표 1-1> 2022년 탄도미사일 도발 일지

구분	횟수[3]	비 고
1월	6	한반도 정세 주도권 확보 조선노동당 정치국 회의 (1.19.)
2월	1	대미 협상력 제고 김정일 생일(2.16.)
3월	3	김정은 국가 우주개발국 현지지도(3.10.)
4월	1	한미 연합 지휘소훈련 (4.18.~26.)
5월	4	윤석열 정부 취임(5.10.) 한미 정상회담(5.21.)
6월	1	한미 해군간 항모강습단 연합훈련(6.2.~4.)
9월	3	한미 연합 해상 훈련 (9.26~29)
10월	6	노동당 창건일(10.10.), 호국 훈련(10.17.~28.)
11월	6	한미 공중 훈련 비질런트 스톱(10.31.~11.5.)
12월	3	유엔총회 북한인권결의안 채택(12.15.)
계	34	-

2022년은 북한이 역대 가장 많은 탄도미사일 도발을 실시한 해로 총 34회를 시행했다[3]. 과거 어떤 시기에 비해 도발 횟수가 매우 많아졌는데, 특히 3월 ~ 5월에 집중되었던 과거와는 다르게[4] 1·10·11월에 6회나 도발을 시행하는 등 다양한 도발 패턴을 보여주었다. 이러한 상황에서는 새로운 도발 패턴에 대한 대비와 더불어 점점 더 발전된 미사일의 위협에 대한 대비가 필요하다. 따라서 적의 새로운 도발 패턴에 대응하고 점점 고도화되는 북한 탄도미사일의 위협에 대응하기 위해 기상청의 기상 데이터를 활용하여 북한 탄도미사일 발사 환경이 기상 환경측면에서 준비되어 있는지를 예측해 보았다. 이를 위해서 과거 북한의 탄도미사일 발사일 기상과 도발을 하지 않은 날의 기상 데이터를 학습데이터로 활용하여 기상적으로 탄도미사일 도발 환경이 준비되어 있는지를 예측하려고 한다.

## 2. 관련 연구

### 2.1 도발 패턴에 관한 연구

북한의 미사일 실험 및 발사 진행 절차를 포괄적 전략(분석틀) 측면에서 보면 주변 정세 파악, 해당 담당자 토의, 실험/발사 계획수립/보고, 김정은 최종결정 순으로 진행된다[4]. 북한은 도발 전 먼저 한국과 미국의 정세를 세부적으로 파악한다[5]. 이후 이해관계를 고려한 실험, 발사 수단 및 방법토의를 시행하는데 이때 관련 내용은 독대로 진행되기 때문에 누구도 관련 내용을 알 수가 없다[4]. 북한이 우리나라와 세계에 가장 위협을 주는 분야가 핵이다. 핵과 관련해서 북한은 핵실험에 앞서 장거리 탄도미사일을 먼저 발사하고 이후 핵실험 예고, 외무성 성명 발표, UN 제재에 대한 반박 성명순으로 반복적인 패턴을 보여 오고 있다. 이러한 북한의 핵실험 관련 도발 패턴을 볼 때 탄도미사일 발사 여부를 예측하는 것은 북한

핵실험에도 대응할 수 있다는 의미가 있다.

### 2.2 탄도 미사일의 발사지점 예측 연구

탐지된 속도를 활용한 발사지점 예측 연구를 보면 발사후 Boost 단계와 Ballistic 단계를 구분하여(PDLS : Phase Division Least Square) 발사지점을 예측한다[6]. 탄도미사일 속도와 각도, 진행방향 중 방위각을 이용하여 Least Square 방법으로 이전 속도를 추정하고, 추정된 속도 특성 및 현재위치와 Euler 적분법을 이용하여 발사지점을 예측하여 최초 탐지 시점이 빠를수록 발사 지점에 대한 거리 오차가 줄어드는 성능을 보여주었다. 하지만 종말단계에서 추력이 없을 때 탄도미사일이 탐지된다면 실제 발사장소와 큰 오차를 보이며 발사를 예측하는 경계의 의미보다 이미 발사된 후 원점을 추정한다는 점에서 한계가 있다. 우리에게 실질적 위협인 북한의 탄도미사일을 대비하기위해선 북한이 발사하기 전 징후를 파악하고 이에 대한 조치가 필요하다.

### 2.3 감시정찰자산 활용 탄도 미사일 발사 탐지 연구

위성 기술을 활용하여 북한의 탄도미사일을 탐지하는 연구를 살펴보면 위성과 지·해상 레이더를 활용한 방안을 제시하고 있다. 그 일환으로 융합센터를 활용한 정지궤도 적외선 위성, 저궤도 SAR 위성, 지·해상 레이더 통합 탐지방안으로 실질적으로 경향을 확인하며 대응하는 방안도 제시되고 있다[7]. 특히 조기경보 위성 기술을 활용하여 북한의 탄도미사일 활동을 감지한다면 탄도미사일 발사 등의 북한위협에 보다 세밀하게 대응할 수 있을 것이다[8]. 하지만 북한은 잠수함이나 터널 등을 이용하며 다양한 발사 수단을 동원하여 탄도미사일을 발사하고 있기 때문에 기존 탐지자산으로 탐지를

더욱 어렵게 하고 있다는 점도 고려해야 한다. 즉, 기존 탐지자산으로는 예측에 한계가 있다.

### 3. 연구 방법

#### 3.1 연구 대상

본 연구의 목적은 북한의 기상 데이터를 활용하여 탄도미사일 발사 가능한 환경을 예측하는 모델을 개발하는 것으로, 실제 북한이 탄도미사일을 발사한 날 기상을 도발할 수 있는 기상으로 가정하고 분석하였다. 기상정보를 확인하려면 지역 정보가 중요한데, 지역 정보는 현재까지 북한이 탄도미사일을 발사할 수 있는 곳으로 예측되는 북한의 9개 도를 대상으로 하였다. 최종적으로 북한 지역 9개 도 중에 비도발지역(북한이 한번도 도발하지 않은 지역)인 양강도, 함경북도, 황해남도를 제외하고 6개 도 지역(강원도, 자강도, 평안남도, 평안북도, 함경남도(2), 황해북도)의 7개 도시(원산, 강계, 평양, 신의주, 신포, 함흥, 개성)를 탄도미사일 도발 가능 지역으로 선정하였다. 선정된 지역에 대한 '22년 기상자료를 수집하고, 탄도미사일과 관련된 관련 연구를 기초로 기상정보를 분석하여 기상정보 가운데 기상변수로 <표 3-1>과 같이 총 11개 변수를 선정하였다.

<표 3-1> 기상 변수

Variable		Unit
avgTa	평균 기온	℃
avgTd	평균 이슬점온도	℃
avgPa	평균 현지기압	hPa
avgWs	평균 풍속	m/s
sumRn	합계 강수량	mm
avgTca	평균 전운량	(0~10)
maxTa	최고 기온	℃
avgHm	평균 습도	%
avgPs	평균 해면기압	hPa
Lmac	중하층 운량	(0~10)
minTa	최저 기온	℃

11개 변수 중 sumRn(결측률 63.78%)과 Lmac(결측률 32.14%)은 다수의 결측치로 인해 모델 구성 간 제외하였다.

기상 변수 중 이슬점온도는 공기가 일정한 압력과 일정한 수증기 함량에서 냉각될 때 포화되는 온도로 이슬이 맺히는 온도이다[9]. 이슬점온도가 낮을수록 공기 중의 수증기 함량이 적어지며 이는 비례관계를 보여준다. 해면기압은 실제 고도가 아닌 평균해면에서 기압으로 환산한 것으로 관측소기압으로부터 경험적으로 추정한다[10]. 해수면 기준으로 기압을 환산했기 때문에 고도에 상관없이 비교할 수 있는 이점이 있다.

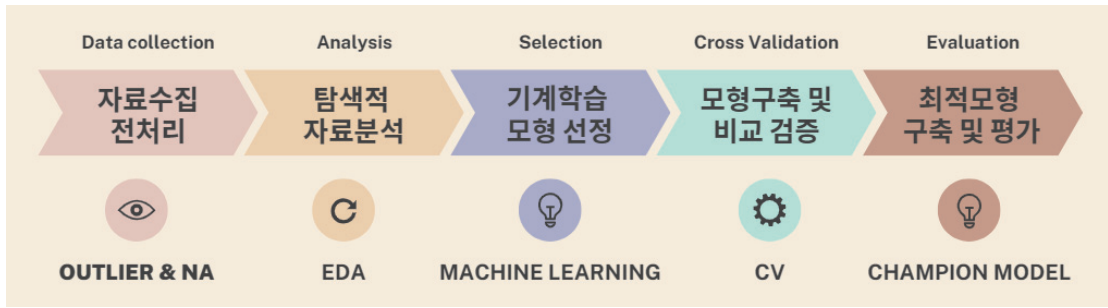
#### 3.2 연구 절차와 방법

북한 탄도미사일 발사 예측 모델을 구성하기 위한 연구 절차는 <그림 3-1>과 같다.

자료수집 및 전처리 단계에서는 북한의 기상 자료를 수집하고 특이값이나 결측치(NA) 등을 확인하는 단계로 자료는 기상청 기상자료개방포털에서 수집하였다<sup>4)</sup>. 1973년부터의 북한기상 데이터가 저장되어있으며 이중 가장 도발이 많이 발생했던 해인 2022년의 자료를 수집하였다. 특이값(outlier)은 각 변수별 상자 그림을 통해서 확인하고 제거하였고, 결측치(missing)를 확인하여 변수 평균으로 대체하였다.

탐색적 자료분석(EDA, explanatory data analysis) 단계는 데이터를 있는 그대로 탐색하여 분석하는 것으로 기술 통계기법 등이 주로 사용된다. 잔차, 자료의 재표현, 그래픽을 통한 현시성을 통해 데이터를 있는 그대로 분석할 수 있다. 상관관계 분석간에는 변수간의 선형적인 관계 정도를 측정하여 그 크기와 방향을 파악할 수 있는데 상관관계가 높다면 다중공선성이 발생할 수 있어 변수를 제거하거나 조정하는 등의 과정이 필요하다.

4) 기상청 기상자료개방포털(<https://data.kma.go.kr/data/grnd/selectNkRltmList.do?pgmNo=58>)



<그림 3-1> 연구 절차 순서도

기계학습 모형 선정은 앞서 준비한 데이터를 다양한 기계학습 모형들 중에 최적의 모형을 구축할 수 있도록 선정하는 단계이다. 모형은 단일모형과 앙상블로 구성하였다. 딥러닝은 제외하였는데 이는 딥러닝에 필요한 만큼의 자료 크기가 충분하지 않았고, 앙상블 학습이 딥러닝만큼의 충분한 성능을 보여주기 때문이다 [12]. 모형 구축 간 K-Nearest Neighbors algorithm(최근접 이웃 알고리즘, 이하 KNN), CART (classification and regression tree) 알고리즘 기반 의사결정나무(이하 Rpart), random forest(이하 RF), support vector machine(이하 SVM), linear discriminant analysis(선형판별분석, 이하 LDA), quadratic discriminant analysis(이하 QDA), Adaboost, generalized linear model(이하 GLM)를 사용하였다. 모형 구축을 위해 R programming 패키지를 이용하였으며 KNN을 위해 ‘class’, LDA와 QDA를 위해 ‘MASS’, SVM을 위해 ‘e1071’, RF를 위해 ‘randomForest’, CART를 위해 ‘rpart’, Adaboost를 위해 ‘ada’패키지를 사용하였다.

GLM(Generalized linear model)은 일반적인 선형 회귀(OLS)를 일반화한 모델로 여기에서 사용한 Logistic Regression은 반응변수가 0(비도발)과 1(도발)인 이진변수일 때 사용한다. SVM은 gamma와 cost 두 개의 초모수(Hyper

parameter)가 있으며, gamma는 커널의 폭(너비)을 조정하는 값이고 cost는 과적합을 막는 정도를 지정한다. 최적의 parameter값을 찾기 위해 tune.svm함수를 사용하였으며 best performance의 gamma는 0.1, cost는 100의 값을 구할 수 있었다.

모형구축 및 비교와 검증단계는 교차타당도 평가를 시행하는 단계로 ‘70대 30’과 ‘10-fold’ 교차타당도 방법을 평가에 사용하였다. 첫 번째로 ‘70대 30’은 훈련자료 70%, 검증자료 30%로 데이터를 분할한 후 무작위로 10회 반복해서 실험하는 방법이며, 두 번째로 ‘10-fold’ 교차타당도 방법은 데이터를 10등분하여 90%로 모델을 학습시키고 나머지 10%로 검증하는데 10등분한 훈련용 90%와 검증용 10%를 바꿔가며 10번 시행하여 그 결과를 평균 내어 검증하는 방법이다. 즉, 각 반복에서 얻은 평가 점수의 평균이 최종적인 모델의 정확도로 추정된다 [11].

최적모형 구축 및 평가는 검증 결과에서 선택된 적합한 모델에 대해 검증에 사용된 모든 데이터와 함께 100%의 데이터를 활용하여 모델을 다시 학습시켜 최적의 모델을 만들었다. 최적의 모델은 OOB(out of bag)를 사용해서 정확도를 평가할 수 있고, 이를 정확도로 하는 최적모형으로 제안하였다[12].

## 4. 연구 결과

### 4.1 탐색적 자료분석

본 연구에서 사용한 데이터는 2022년 북한의 기상 제원이다. 평균 기온(avgTa), 평균 풍속(avgWs) 등 9개 연속형변수 기술통계량을 비교한 결과, avgHm과 avgPa, avgTd 분산이 각각 264.18, 196.65, 166.39로 측정되었다. 이러한 결과는 해당 변수들의 연간 변화율이 상대적으로 크다는 북한 기상의 특징을 보여준다.

연속형 변수간 선형적인 상관관계를 알아보기 위해 피어슨 상관계수를 산출하였다. avgTa

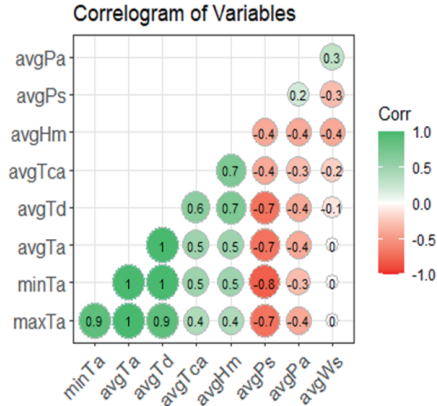
와 avgTd, minTa 등 온도는 상관계수가 1로 강한 선형성을 보여주었다. avgPs와 minTa는 -0.8로 강한 음의 상관관계를 보였고 avgPs와 avgTa, avgTd, maxTa 등 평균 해면기압과 기온간에는 음의 상관관계를 보여주었다. 이는 평균 해면기압이 클수록 기온은 낮아지는 것을 보여주는데 해면기압과 기온이 반비례 관계라는 현상과도 일치한다[13]. 즉, 해수면 공기의 온도가 상승하면 공기가 팽창하기 때문에 밀도가 낮아져 기압이 낮아지게 된다. 반대로 온도가 낮아지면 공기가 수축하게 되며 밀도가 높아져 압력이 증가한다.

<표 4-1> 연속형 변수 기술통계량

변수(단위)	최소	1/4분위수	중앙	평균	3/4분위수	최대	분산
avgTa(°C)	-18.4	1.9	12.2	10.71	20.4	30.5	123.81
avgTd(°C)	-23.5	-5.3	5.3	4.67	15.8	26.4	166.39
avgPa(hPa)	959.6	1001.6	1010.6	1007.8	1018.0	1035.6	196.65
avgWs(m/s)	0	0.8	1.6	1.7	2.4	9	1.46
avgTca(0~10)	0	1.5	5.1	5.01	8.4	10	12.71
maxTa(°C)	-15.9	7.25	18.1	16.43	25.9	37.6	121.39
avgHm(%)	22	58	71	69.19	82	99	264.18
avgPs(hPa)	991.2	1010.8	1018.0	1017.7	1024.6	1043.4	85.35
minTa(°C)	-25.5	-3	7	6.02	16.1	25.9	138.27

<표 4-2> 기상변수간 상관관계 분석

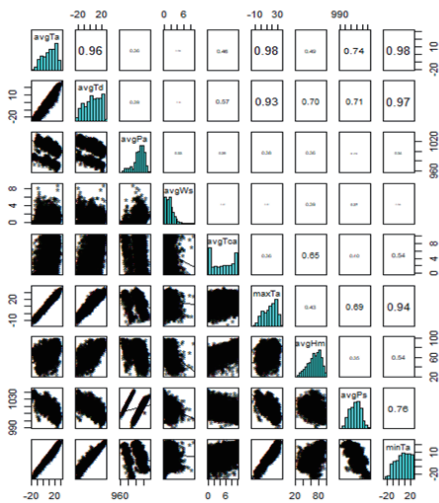
변수	avgTa	avgTd	avgPa	avgWs	avgTca	maxTa	avgHm	avgPs	minTa
avgTa	1	-	-	-	-	-	-	-	-
avgTd	1	1	-	-	-	-	-	-	-
avgPa	-0.4	-0.4	1	-	-	-	-	-	-
avgWs	0	-0.1	0.3	1	-	-	-	-	-
avgTca	0.5	0.6	-0.3	-0.2	1	-	-	-	-
maxTa	1	0.9	-0.4	0	0.4	1	-	-	-
avgHm	0.5	0.7	-0.4	-0.4	0.7	0.4	1	-	-
avgPs	-0.7	-0.7	0.2	-0.3	-0.4	-0.7	-0.4	1	-
minTa	1	1	-0.3	0	0.5	0.9	0.5	-0.8	1



<그림 4-1> 기상변수간 상관관계 분석(1)

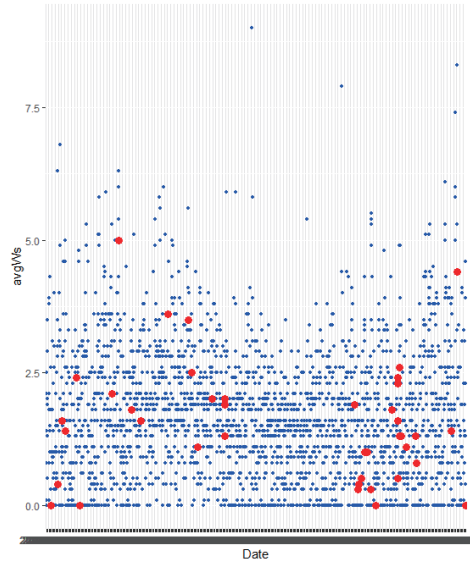
<그림 4-1>과 <그림 4-2>의 상관관계 분석 간 산점도 그래프를 보면 avgTd와 avgTa, avgHm, maxTa, minTa간 강한 선형성을 보여 주고 있다. 이는 평균 이슬점온도가 기온에 영향을 받는다는 것을 유추해볼 수 있다. 기온이 높을수록 대기 중의 수증기 함량이 높아지기 때문에 평균 이슬점 온도도 높아지며, 기온이 낮을수록 대기 중의 수증기 함량이 낮아져 평균 이슬점 온도도 낮아진다. 따라서 이러한 그래프 결과는 평균 이슬점 온도와 기온 간에 밀접한 관계가 있음을 보여준다.

scatter-plot matrix, correlation coef., histogram

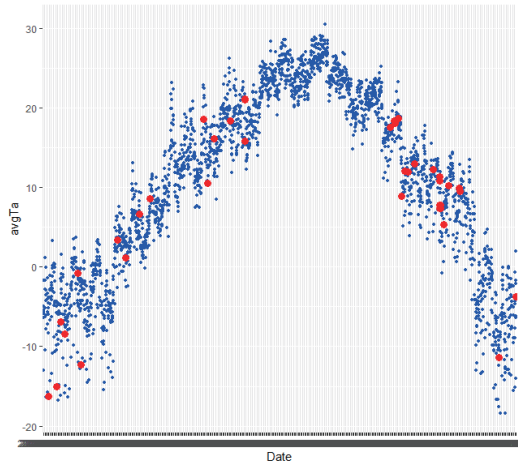


<그림 4-2> 기상변수간 상관관계 분석(2)

<그림 4-3> avgWs 일자별 도발분포 그래프는 연간 일별 평균 풍속을 시각화한 것이다. 파란 점은 비도발 지역이고, 빨간 점은 도발 지역으로 탄도미사일은 대체로 풍속이 0~2.5m/s일 때 발사되었음을 알 수 있다.



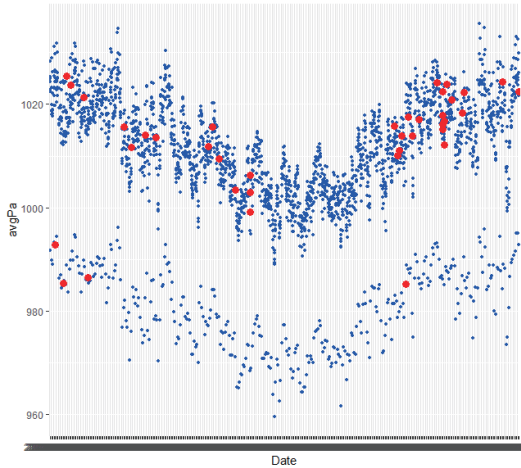
<그림 4-3> avgWs 일자별 도발분포



<그림 4-4> avgTa 일자별 도발분포

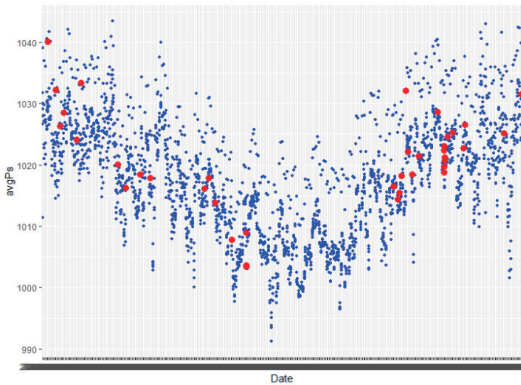
<그림 4-4> avgTa 평균 기온에 따른 발사일을 분석해보면 영하 15도부터 20도까지 발사한 분포를 보여준다. 평균 기온 20도 이상일

때 발사횟수는 1회에 불과하다. 이러한 결과는 기온이 탄도 미사일 발사에 영향을 줄 수 있는 것으로 볼 수도 있으나 과거 사례를 분석해볼 때(2012~2017년 김정은시기 7월 7회, 8월 3회 도발<sup>4)</sup>) 다양한 변수에 대한 고려가 필요하다.



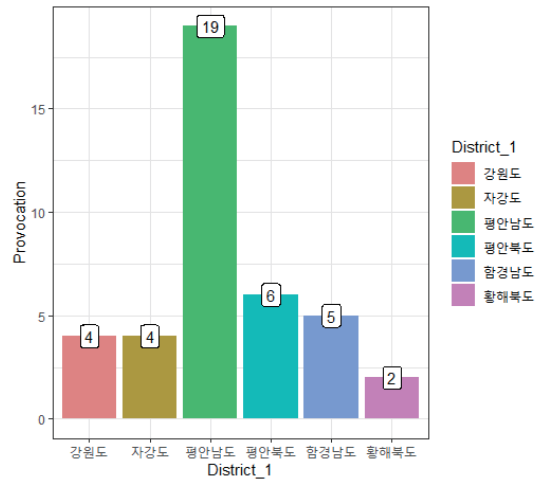
<그림 4-5> avgPa 일자별 도발분포

<그림 4-5>에서 아래 평균현지기압이 낮은(960hpa ~ 995hpa) 지역은 북한 자강도 강계시 지역으로 대체적으로 낮은 현지기압을 보여주었다. 북한 자강도 해발고도가 타 지역보다 높음을 알 수 있는데 평균고도는 해발 803m<sup>5)</sup>로 비교적 높은 편에 속한다.



<그림 4-6> avgPs 일자별 도발분포

<그림 4-6> avgPs 평균해면기압은 어떤 관측지점에서 관측된 현지기압을 평균 해수면(해발고도 0m)기준으로 환산하여 산출한 기압이다. 해면기압은 고도가 달라도 비교 가능하여 기상청에서 일기도를 그리는 데 사용된다. 고기압이 형성되는 지역은 대기가 하강하여 공기가 밀집하게 쌓이는 지역이기 때문에 해면기압이 상승하게 된다. 7개 도시 모두 비슷한 형태를 보여주며 대체로 겨울엔 높았으며 여름에는 공기가 뜨거워져서 상승하게 되므로 해면기압이 낮아지는 것을 보여준다.



<그림 4-7> '22년 도발지역별 횟수

도발지역별 도발횟수를 시각화해서 보면 <그림 4-7>과 같이 평안남도(평양 포함)에서 가장 많은 도발이 이루어졌음을 알 수 있다. 주로 순안구역 비행장 일대에서 동해상으로 발사하였으며(15회) 그 외 평안남도 4곳에서 탄도미사일을 발사하였다. 발사체 또한 다양화되고 있는데 평안북도에선 일부 철도 발사형으로 발사하였으며 함경남도 신포 앞바다에선 잠수함에서 발사하는 등 점점 더 예측이 어려워지고 있다.

5) 한국민족문화 대백과사전(<https://encykorea.aks.ac.kr/Article/E0047852>)

## 4.2 데이터 불균형 해소

SMOTE(Synthetic Minority Over-sampling Technique)는 머신러닝에서 불균형한 데이터를 처리하는 오버샘플링 방법의 하나다. 이 방법은 소수 클래스의 데이터를 무작위로 선택한 후, 그 데이터와 가장 가까운 이웃 데이터들 사이에 새로운 데이터를 합성하는 과정을 반복하여 다수 클래스의 데이터와 균형을 맞춰준다. 이 방법은 합성 데이터를 생성함으로써 샘플의 크기를 증가시키는데, 이를 통해 모델이 불균형 데이터를 처리할 수 있도록 도와준다. 본 연구에서는 1단계로 도발일과 비도발일의 기상을 분석하여 균형을 맞추고 2단계로 SMOTE방법을 적용하였다.

1단계로 도발일과 비도발일의 데이터 불균형이 크기 때문에 불균형데이터를 균형있게 맞춰주는 과정을 거쳤다. 기상으로 탄도미사일 발사 가능한 환경을 예측하기 위한 목적상 각 기상변수의 도발일 기상 평균  $\pm 1\sigma$ (표준편차) 내에 있으면 도발 가능한 것으로 보고 도발일로 가정하였다. 즉, 도발일에 도발하지 않은 지역 이더라도 기상변수별 도발일 기상 평균  $\pm 1\sigma$  내에 있다면 도발이 가능하다고 보았다.

<표 4-3> Class 수 비교

Category	N-Provocation	Provocation
1단계 (기상 비교 분석)	2493	62
2단계 (SMOTE)	2493	2480

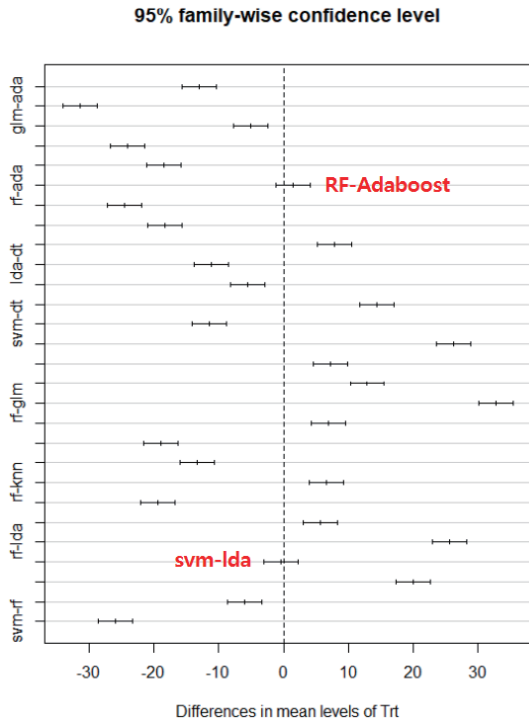
2단계로 SMOTE 방법을 적용하기 전에는 비도발일과 도발일의 데이터 수의 차이가 매우 컸지만, <표 4-3>과 같이 SMOTE를 통해 불균형의 문제를 보완하였다.

## 4.3 모형구축 및 비교검증

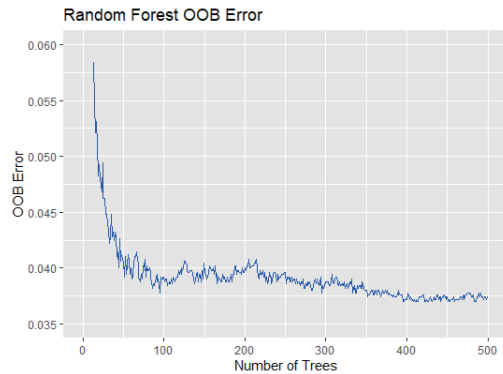
앞서 제시한 8개의 모델로 '70대 30'과 '10-fold' 교차타당도 평가를 하였다. <표4-4> 교차타당도 평가 결과를 보면 'RF'와 'Adaboost'가 가장 좋은 정확도를 보여준다. 결과가 유의미한지 확인하기 위해 10-fold CV결과를 분산분석한 결과 'F(7,72) = 418.1'이고 유의확률(p-value)이  $< 2e-16$ 으로 유의미한 결과를 보여주었다. 즉, 모형간에 정확도에 대한 차이가 있었다. 분산분석 만으로는 어떤 모델결과간에 유의미한 차이가 있는지 알 수 없기 때문에 사후검정을 실시하였다. 이는 분산분석이 집단간의 차이유무만을 말할 뿐 어느 집단간에 차이가 있는지는 알 수 없으며 사후검정을 통해 어느 집단간에 유의미한 차이가 있는지를 확인하기 위해서이다. Tukey의 다중비교를 한 결과 <그림 4-8>과 같이 'RF'와 'Adaboost'는 서로를 제외한 다른 6개 모델과 통계적으로 95% 신뢰수준 하에서 유의미한 차이가 있었다. 'RF'와 'Adaboost'는 Tukey 다중비교 결과 유의확률(p-value)이 0.71로 차이가 있다고 보기 어려웠다.

<표 4-4> 교차타당도 평가 결과

Category		GLM	KNN	Rpart	RF	SVM	LDA	QDA	Adaboost
70:30	평균 (%)	63.41	89.18	82.49	<b>95.58</b>	70.65	71.04	76.34	<b>94.34</b>
	표준편차	1.13	0.93	2.05	0.57	0.99	1.00	1.04	0.67
10-fold	평균 (%)	63.54	89.80	81.92	<b>96.34</b>	70.40	70.82	76.43	<b>94.93</b>
	표준편차	1.52	2.13	2.69	1.08	2.19	2.20	1.69	0.94



<그림 4-8> 모형들의 Tukey 다중비교 결과



<그림 4-9> Random Forest OOB Error

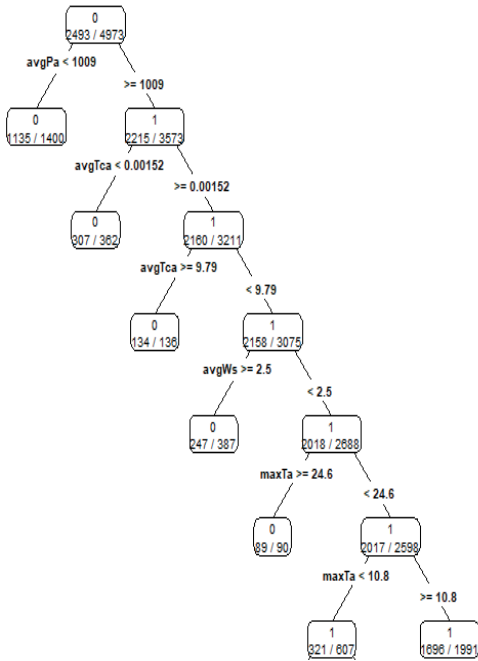
### 4.4 최적모형 구축 및 평가

교차타당도 평가결과 높은 정확도 및 다른 모델과 비교해서 유의미한 차이를 보여준 앙상블모형인 ‘RF’와 ‘Adaboost’ 모델이 최적의 모형이었다. 해당 모델에 대해 검증데이터를 포함한 100% 데이터를 사용하여 최적모형을 구축하였으며 ‘RF’는 96.26%의 높은 정확도를 보여주었다.

‘RF’의 hyper parameter인 Tree의 수를 구하기 위해 OOB Error를 비교한 결과 400개부터 성능 개선이 미비한 것을 알 수 있다. 트리의 수는 알고리즘의 예측 성능과 모델의 복잡도를 조절하는 중요한 hyper parameter이다. 즉, 높은 예측 성능을 보여주면서 모델의 복잡도가 증가하지 않게 적절한 수의 트리를 설정하는 것이 중요하다. 따라서 적정한 트리의 수는 400으로 결정할 수 있었다

Adaboost는 96.0%의 분류 정확도를 보여주었다. ‘RF’와 0.26% 정확도 차이를 보여주었으나 Tukey 다중비교간 유의미한 차이가 없었으므로 ‘RF’ 성능이 더 좋다고 단정하기 어렵다. 이러한 모델을 통해 과거 기상 데이터를 바탕으로 학습하고 새로운 데이터를 입력하여 탄도 미사일 발사 가능한 기상 환경을 예측하는데 활용할 수 있을 것이다.

### 4.5 의사결정나무활용 시사점 분석



<그림 4-10> rpart를 활용한 의사결정나무 모형 분류 결과

rpart를 활용한 의사결정나무 모형 분류 결과는 <그림 4-10>과 같다. 주요 분류변수를 해석해보면 첫 번째로, 현지기압이 1009보다 높고 평균전운량이 0.00152~9.79사이일 때 평균풍속이 2.5보다 크거나 같으면 도발이 불가능한 환경이 63.8%(247/387)이었다. 풍속은 실제로 로켓 발사와도 관계가 깊다. 2022년 6월 14일 누리호 2차 발사가 강풍에 의해 연기되었던 것처럼 지상풍이 너무 강하면 발사체 이동 및 설치, 발사 간 위험이 크기 때문에 로켓 발사에 부정적인 영향을 준다고 볼 수 있다.

두 번째로, 첫 번째 환경에서 평균 풍속이 2.5보다 작고 최고 기온이 10.8~24.6도일 때 85.2% (1696/1991)의 탄도미사일 도발 가능한 환경을 보여주었다. 온도가 발사체의 설계 및 운용에 미치는 영향을 보면 연료탱크 등의 밀폐 용기의 내부 압력 변화, 가열로 인한 열 스

트레스, 발사체의 표면 및 내부의 온도 상승으로 인한 기기의 영향등 다양하며 이에 대한 고려 또한 필요하다[14].

### 5. 결론

본 연구에서는 북한 탄도미사일 발사 가능 기상 환경을 예측하기 위해 북한 기상자료를 활용한 예측 모델을 제시하였다. 본 연구에서는 ‘RF’가 96.34%, ‘Adaboost’가 94.93%의 높은 정확도를 보여주었다.

또한, 탄도미사일 발사 가능 기상 환경 예측에 중요하게 영향을 주는 기상 변수들을 의사결정나무 모형의 결과를 이용해서 해석해 보았다. 해석 결과, 현지기압이 1009보다 높고 평균전운량이 0.00152~9.79 사이면서 풍속이 2.5보다 크거나 같을 때 탄도미사일 발사에 불리한 환경을 보여주었다. 이는 실제 지상풍이 너무 강하면 로켓 발사에 부정적인 영향을 주는 것과 동일한 결과를 보여준다. 반대로 풍속이 2.5보다 작고 최고 기온이 10.8~24.6도일 때 85.2%의 도발 가능한 환경을 보여주었다.

본 연구에서 제안한 탄도미사일 발사 가능 기상환경 예측 모델은 기상 데이터만을 활용하여 분석하였는데, 여러 한계점이 존재한다. 특히, 2022년 북한 기상 데이터만을 사용한 것이기 때문에 다양한 변수들을 추가적으로 수집하여 분석하는 것이 향후에 지속될 필요가 있다.

북한의 탄도미사일 개발은 점점 고도화되고 있으며, 우리나라의 안보에 큰 위협이 되고 있다. 따라서 이를 예측하고 대응하는 노력이 필요하다. 앞으로 본 연구에서 제시한 기상에 대한 분석뿐만 아니라 국·내외 정치적, 군사 외교 및 상황 등과도 연계한 의도 분석이 함께 이루어져야 한다. 이를 통해 우리나라의 안보를 지키고 국민을 보호하는데 기여할 수 있을 것이다.

## 참 고 문 헌

- [1] 박창권, “북한의 탄도미사일 위협과 한국의 대응체제 발전방향”, 국방정책연구 28(2), 2012
- [2] Yong-Sup Han, “Going Into the ‘unification tunnel’: Strategic and military implications“ in Emerging Threats, Force, Structure, and the Role of Air Power in Korea, edited by Natalie W. Crawford and Chung-In Moon, RAND, 2000.
- [3] 국방부, 「2022 국방백서」, 2022.
- [4] 김경목, 백연주, “북한의 미사일 발사 및 핵실험 패턴에 관한 연구”, 전략연구, 73, 2017.
- [5] 홍민, “김정은 정권 핵·미사일 활동의 주요 특징과 패턴”, 통일연구원 Online Series, 2017
- [6] 김준기, 이동관, 조길석, 송택렬, “단계 분리형 최소 자승법을 이용한 탄도 미사일의 발사지점 예측 연구”, 제어로봇시스템학회, 20(4), 2014
- [7] 김수임, “위성과 지·해상 레이더를 활용한 탄도 미사일 통합 탐지방안 제안 및 효과 분석”, 2021, 아주대학교 석사논문
- [8] 백명진, 최성봉, “한국형 조기경보위성 시스템”, 제어로봇시스템학회, 2012
- [9] 기상청 기상용어사전, Available: <https://data.kma.go.kr/cmmn/dictionaryPopup3.do?wordId=5910>
- [10] 기상청 기상용어사전, Available:<https://data.kma.go.kr/cmmn/dictionaryPopup3.do?wordId=1663>
- [11] 문호석, “기계학습을 이용한 위게임 모델의 근접전투 전장상황 평가를 위한 전문가시스템 연구”, 한국군사학논집, 74(3), 2018.
- [12] 박유성, 『파이썬을 이용한 통계적 머신러닝』, 자유아카데미, 2022.
- [13] 류시찬, “동아시아 해면기압과 한반도 겨울철 기온의 관계”, 2004, 강릉대학교 석사논문
- [14] 김준, 국방재, 문경정, 이재화, 구자호, 박상서, 이효근, “외나로도 우주센터의 기상특성”, 한국항공우주연구원, 2009.
- [15] 김은영, “북한 미사일 주요 일지 및 발표문”. 북한경제리뷰, 3, 2009.
- [16] 정규수, 『ICBM 그리고 한반도』, 지성사, 2012.
- [17] 정규수, 『ICBM, 악마의 유혹』, 지성사, 2012.

## 저자 소개



**조성일(E-mail: enjoy\_all@naver.com)**

2016 육군3사관학교 전자공학 학사  
현재 국방대학교 군사운영분석전공 석사과정  
관심분야 : 국방 빅데이터·AI, 텍스트 마이닝



**문호석(E-mail: hsmoon0329@kndu.ac.kr)**

1994 육군사관학교 화학과 학사  
2003 고려대학교 전자공학 석사  
2006 고려대학교 산업공학 박사  
2010 고려대학교 통계학 박사  
현재 국방대학교 국방과학학과 교수  
국가안전보장문제연구소  
군사과학연구센터장  
관심분야 : KCTC 빅데이터분석, 데이터과학,  
국방 AI, M&S



**이춘주(E-mail: sarang90@korea.kr)**

1991 육군사관학교 졸업(이학사)  
1997 미국 UC Berkeley 핵공학 석사  
2005 KDI 국제정책대학원 정책학 석사  
2006 서울대학교 기술정책학 박사  
현재 국방대학교 무기체계전공 교수  
관심분야 : 기술정책, 국방로봇(무인전력),  
방위산업정책, 생산성분석,  
핵심인프라방호



## AHP(Analytic Hierarchy Process)를 활용한 프로젝트 소과제별 중요도 평가 방법

### Method of evaluating the importance of tasks in a project using AHP(Analytic Hierarchy Process)

강승욱<sup>1)</sup> · 윤진성<sup>1)</sup> · 소준호<sup>2)</sup> · 이용복<sup>3)</sup>

Seunguk Kang · Jinseong Yun · Junho So · Yongbok Lee

#### ABSTRACT

Today, Korea military is establishing various initiatives that will lead the transformation of the future forces. Specifically, for plans involving tasks of different characteristics, grouping these tasks and applying scientific evaluation tools is critically important to maintain the directionality of the plan. Therefore, this paper has researched a methodology for evaluating the importance of tasks in a project using the Analytic Hierarchy Process (AHP) as part of a methodology for an objective and comprehensive analysis and evaluation of the plan. Through AHP, we structured tasks and presented the process of deriving task priorities (weights) through surveys as a research. The importance of each task derived from the results of this study will play a very important role in evaluating the performance of the military's crucial implementation plans by being utilized to quantify progress in the future. Through this, it is expected that the implementation plan will maintain the right direction and momentum, contributing to the development of the military's combat power.

Keywords : AHP(Analytic Hierarchy Process), project evaluation, importance of tasks

---

논문접수일 : 2023년 5월 31일, 심사일 : 2023년 5월 31일, 게재확정일 : 2023년 6월 8일

1) 국방대학교 국방과학학과 석사과정

2) 국방부 국방개혁실 군구조혁신분석담당

3) 국방대학교 국방과학학과 교수 / 교신저자(miliman@naver.com)

## 1. 서론

오늘날 우리 군은 북 핵·미사일 위협, 미·중 패권 경쟁, 전쟁 패러다임의 변화, 미래 인구절벽 등 많은 도전요인에 직면하고 있다. 국방부는 이러한 도전요인을 극복하고 국방개혁 2.0의 한계를 획기적으로 극복하기 위해 『국방혁신 4.0』을 발표하였고, 이를 통해 싸워 이기는 AI 과학기술 강군을 육성하는 것을 목표로 국방의 혁신적 변화를 추구하고 있다[1].

국방혁신 4.0에서 군별로 미래 군의 변화를 이끌어갈 다양한 추진계획들이 수립되고 있는 상황 속에서 추진계획의 목표와 최종상태 달성을 위해서는 과제의 객관적인 점검·평가를 통해 추진동력과 방향성을 유지하는 것이 중요하다.

하지만 지난 국방개혁 2.0의 사례에서도 알 수 있듯이 현재 국방부의 평가체계는 매우 미흡한 실정이다. 국방개혁 2.0 추진 당시 분기·반기별로 국방개혁 2.0 추진평가 회의를 시행했는데, 평가 간에 과제 담당 또는 추진 시행부서에서 자체 평가 후 진도를 추정하는 방식으로 진행되었다. 이러한 방식은 객관성 및 신뢰성이 보장되지 않고, 과제별 평가 기준 또한 모호한 점이 있었다[2].

또한, 정량적 과업과 정성적 과업으로 분류되는 국방개혁과 같이 다른 여러 과제가 혼합된 추진계획에 대하여 객관적이고 종합적으로 분석평가 할 수 있는 방법론 연구는 부족한 현실이다.

본 논문은 AHP를 활용하여 프로젝트 소과제별 중요도를 평가하는 방법론을 국방혁신 4.0 과제를 중심으로 연구하였다. 추진계획에 대한 과학적이고 종합적인 분석평가를 위해서는 평가 항목과 요소 선정부터 평가유형 그룹화, 평가구조 개발까지 다양한 방법론이 융합되어야 한다. 본 연구는 이 과정에서 가장 중요한 프로

젝트 소과제별 중요도를 평가할 방법을 제시하였다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 국방개혁 평가와 관련된 연구와 AHP 기법의 이론과 일반적인 내용을 살펴본다. 3장에서는 국방혁신 4.0의 개념, AHP 계층구조와 설문 종합 결과를 설명하고, 4장에서는 결론 및 향후 연구 방향을 제시하였다.

## 2. 이론적 배경

### 2.1 국방개혁 평가 관련 기존 연구

국방개혁 기본계획은 『국방개혁 2020』이 2005년에 발표된 이래로 2018년에 『국방개혁 2.0』이 발표되기 이전까지 3차례의 수정과 보완을 거쳤으며, 이 과정에서 국방개혁에 대한 평가에 관련된 다양한 연구가 진행되었다[3][4][5]. 김상범(2006)은 국방개혁 2020의 추진 방향을 공중전력을 중심으로 외국의 국방개혁과 비교 분석하였고, 이를 토대로 발전과제와 방향을 제시하였다[6][7]. 형혁규(2020)는 국방개혁 2.0이 진행되고 있는 시점에 연구를 진행하였는데 역대 국방개혁 기본계획 내용을 전체적으로 비교하고, 국방개혁 2.0의 주요 성과를 평가하고 앞으로 추진해 나가야 할 과제를 제시했다[3]. 이상철(2021)은 국방개혁 2.0의 성과를 진단하고 발전 방향을 제시하는 연구를 진행하였는데, 특히 국방개혁 추진에 대한 객관적인 평가체계의 부재를 미흡점으로 도출하여 평가체계를 구축하는 것이 필요함을 밝혔다[2].

결론적으로 국방개혁과 관련된 다양한 연구가 진행되었지만, 추진계획을 객관적이면서 종합적으로 분석 평가할 수 있는 방법론 연구는 부족하다.

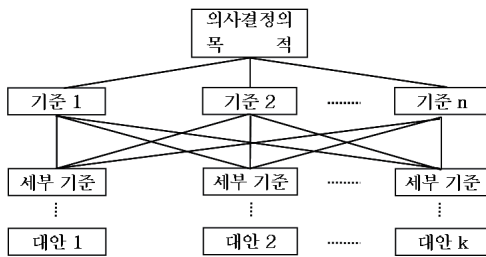
### 2.2 AHP 이론 소개

AHP(Analytic Hierarchy Process, 계층분석적 의사결정 방법)는 의사결정의 계층구조를 구성하고 있는 요소 간의 쌍대비교를 통하여 평가자의

지식, 직관 등을 포착하고자 하는 하나의 의사결정 방법이다[8]. AHP는 이론의 단순성 및 명확성, 적용의 간편성, 범용성의 장점으로 인해 여러 의사결정 분야에서 널리 응용되어 왔다. AHP는 복잡한 상황의 구조화, 비율척도를 통한 우선순위(가중치) 도출, 통합 및 논리적 일관성 검증 등이 가능하다.

AHP를 사용할 때는 일반적으로 다음과 같은 단계의 작업이 수행된다.

첫째, 의사결정 문제를 상호 관련된 의사결정 사항들의 계층으로 분류하여 의사결정 계층을 설정한다. AHP 방법에서 가장 중요한 단계라고 할 수 있는데, 계층의 최상층에는 가장 포괄적인 의사결정의 목적이 놓이며 계층이 낮아질수록 요소가 구체적인 것이 된다. 이러한 의사결정 체계는 <그림 2-1>과 같이 나타난다.



<그림 2-1> AHP의 표준 계층<sup>4)</sup>

$$A = \begin{bmatrix} 1 & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & 1 & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & 1 & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & 1 \end{bmatrix}$$

<그림 2-2> 쌍대비교행렬

둘째, 의사결정 요소 간의 쌍대비교로 판단자료를 수집한다. 이 단계에서는 상위계층에 있는 요소들의 목표를 달성하는데 공헌하는 직계 하위계층에 있는 요소들을 쌍대비교하여 행렬을 작성한

다. 통상 설문을 활용하여 쌍대비교를 통해 상위 요소에 기여하는 정도를 종합하게 되는데, 9점 척도로 중요도를 부여하게 된다. 작성된 쌍대비교행렬 A는 <그림 2-2>와 같이 행렬의 대각을 중심으로 역수의 형태를 취하게 된다. ( $a_{ji} = 1/a_{ij}$ ,  $a_{ii} = 1$ )

셋째, 쌍대비교행렬이 종합되면 고유값 방법을 사용하여 의사결정 요소들의 상대적인 가중치를 추정한다. 한 계층 내에서 비교 대상이 되는 n개 요소의 상대적인 중요도를  $w_i (i = 1, \dots, n)$ 라 하면, 쌍대비교행렬 A에서의  $a_{ij}$ 는  $w_i/w_j (i, j = 1, \dots, n)$ 로 추정할 수 있다. 추정된  $a_{ij}$ 를 이용하여 행렬 A를 나타내게 되면 고유값 방법에 의하여 식 (1)에 의해 w를 구할 수 있다[8].

$$A \cdot w = n \cdot w \tag{1}$$

$w = [w_1 \ w_2 \ w_3 \ \dots \ w_n]$  : 행렬 A의 고유벡터

n : 행렬 A의 고유값

이 과정에서 정확한 w를 구할 수는 없다. 따라서 쌍대비교행렬 A의 각 요소에 대한 가중치 w를 모른다고 했을 때, 이 행렬을 A'라 하고 다음과 같은 식 (2)로 w'를 추정한다.

$$A' \cdot w' = \lambda_{\max} \cdot w' \tag{2}$$

$\lambda_{\max}$  : 행렬 A'의 가장 큰 고유값

이렇게 구한 고유값을 통하여 평가 결과에 대한 일관성 검증을 하게 되는데, 일관성 지수는 식 (3)과 같이 구하고 경험 법칙에 따라 값이 0.1보다 작으면 쌍대비교행렬은 일관성이 있다고 판단한다.

$$CI = (\lambda_{\max} - n) / (n - 1) \tag{3}$$

4) 이용복, “서열척도를 이용한 AHP 그룹의사결정 방법 - 무기체계 획득 과정을 중심으로”, 한양대학교 석사학위논문, 2006

넷째, 일관성 지수가 기준에 부합한 것으로 확인이 되면, 평가대상이 되는 대안들에 대한 종합 순위를 얻기 위해 의사결정 요소들의 상대적인 가중치를 종합한다. 이때 본 연구를 비롯하여 많은 현실 문제들은 다수의 평가자가 참여한 그룹 의사결정의 형태로 이루어지며, 그중에서도 ‘수치 통합방법’을 통해 그룹의 평가치를 주로 종합한다. 이는 그룹 멤버가 행한 각각의 쌍대비교행렬을 수집하고 그룹 전체의 평가치를 수치 통합하여 가중치를 구하는 방법이다. 수치 통합방법은 세 가지의 방법으로 분류되나 특별한 경우를 제외하고는 기하평균을 이용한 방법을 사용하는 것이 바람직하므로, 해당 방법을 소개한다.[8] 전체 평가자가  $n$ 명으로 구성되며  $a_{ij}$ 를  $k$ 번째 평가자가 평가한 쌍대비교행렬에서의 각 원소라 할 때, 통합된 단일 쌍대비교행렬의 각 원소는 식 (4)와 같다.

$$= \prod_{k=1}^n (a_{ijk})^{1/n} \quad (4)$$

기하평균을 사용하는 이유는 행렬의 역수성을 유지시키는 유일한 방법이며, 이상치의 영향을 너무 크게 받는다는 산술평균의 단점을 보완할 수 있기 때문이다[9].

### 3. 연구 방법

#### 3.1 국방혁신 4.0 개념

국방환경은 북 핵·미사일 위협의 현실화와 고도화, 미·중 패권 경쟁에 따른 동북아 불안정성 심화, 전쟁 패러다임의 변화와 기술패권 경쟁 격화, 인구절벽에 따른 병역자원 감소와 같은 도전요인에 직면해있다. 국방혁신 4.0은 국방환경 도전요인을 AI 과학기술 강군 육성을 통해 북한 핵·미사일 위협에 대한 대응능력을 획기적으로 강화함으로써 대북 억제를 달성하고

AI·무인·로봇 등 첨단과학 기술 기반 경쟁우위의 작전 수행능력을 갖추는 데 목적이 있다. 이를 위해 AI·무인·로봇 등 4차 산업혁명 과학기술 기반으로 북 핵·미사일 대응, 군사전략 및 작전개념, 핵심 첨단전력, 군구조 및 교육훈련, 국방 R&D·전력증강체계 분야를 혁신하여 경쟁우위의 AI 과학기술 강군을 육성하는 것이 핵심 추진전략이다. 국방개혁 2.0과 국방혁신 4.0의 차이점은 <그림 3-1>과 같으며 주요 내용은 다음과 같다. 국방개혁 2.0은 북한 위협이 점진적으로 감소하고 있다는 인식하에 국방 전 분야를 망라하여 단기적 관점에서 양·규모 축소에 중점을 두었다. 이와 달리 국방혁신 4.0은 북 핵·미사일 위협이 현실화되고 있음을 명확히 인식하고 첨단과학기술 관련 핵심분야에 집중하여, 중·장기적 추진 관점에서 국방 전 분야의 질적 향상을 추구한다는 점에서 차이가 있다고 할 수 있겠다[10][11][12][13].



<그림 3-1> 「국방개혁 2.0」과 「국방혁신 4.0」의 차이점

### 3.2 국방혁신 4.0 과제별 중요도 평가의 필요성

국방혁신 4.0은 AI 과학기술 강군 육성을 목표로 5개의 중점을 가지고 추진되고 있다. 추진 중점에 관한 내용을 살펴보면 현실적인 위협인 고도화되는 북한 핵·미사일 위협에 대한 ‘북핵·미사일 대응능력 획기적 강화’, 미래 안보 환경에 부합하는 ‘군사전략·작전개념 선도적 발전’, 미래 합동 작전개념을 구현하고 미래 전장을 주도하기 위한 ‘AI 기반 핵심 첨단전력 확보’, 미래 전장 환경에 최적화된 군 구조를 마련하고, 전투력을 극대화하기 위한 ‘군구조 및 교육 훈련 혁신’, 민간 우위의 첨단 기술을 국방에 접목하고 전력화 시기를 단축하기 위한 ‘국방 R&D·전력증강 체계 재설계’ 이렇게 구성되어 있다. 이러한 추진 중점 아래 중점별 2~4개 과제를 구성하여 단계별로 추진하고 있다. 중점별 과제는 <그림 3-2>와 같다.



<그림 3-2> 「국방혁신 4.0」 중점별 과제

중점별 과제는 다수의 과업으로 구성되며, 각 과업을 사업추진방식에 따라 정량적 유형과 정성적 유형으로 구분된다. 또한, 과업별 추진 시

기, 예산 요구사항, 국방혁신 4.0의 최종목표 달성에 미치는 중요도들이 다르다. 따라서 국방혁신 4.0 과제별 중요도를 평가하는 것은 반드시 필요하다고 볼 수 있다.

### 3.3 가중치 결정을 위한 계층구조 및 설문 작성

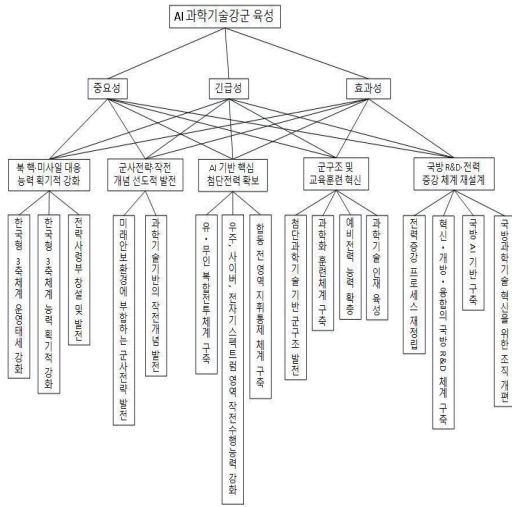
다른 유형과 시기, 중요도를 갖는 과제를 일괄적으로 종합하여 전체 성과를 평가하는 것은 비현실적이기 때문에 각 과제에 대한 가중치를 다르게 부여하여 전체 성과를 종합하는 것이 필요하다. 각 분야와 각 과제에 대한 가중치를 부여하기 위해서는 일관된 평가 기준(Criteria)이 필요하다.

평가 기준은 국방환경 도전요인을 해소하고, 국방혁신 4.0 목표 및 최종상태 달성을 위한 추진성과를 평가할 수 있는 항목으로 선정하였다.

먼저 시간 관리 분야의 고전인 아이젠하워 원칙으로부터 국방혁신 4.0의 최종상태 달성과 ‘AI 과학기술강군 육성’ 목표달성에 중요한 정도를 나타내는 ‘중요성’과 국방환경 도전요인의 해소와 국방혁신 4.0 목표 및 최종상태 달성을 위해 우선 추진해야 하는 정도를 나타내는 ‘긴급성’ 항목 선정하였다. 또한, 국방혁신 4.0 분야 및 과제별 혁신결과가 국방환경 도전요인을 해소하여 국방혁신 4.0 목표와 최종상태 달성에 기여하는 효과를 평가할 수 있는 ‘효과성’ 항목을 추가로 선정하였다. 그 아래 계층으로 국방혁신 4.0의 5대 추진 중점과 16개 과제를 구성하였다. 총 3개의 계층으로 구성하였으며 국방혁신 4.0 과제별 중요도 평가를 위한 세부적인 계층구조는 <그림 3-3>과 같다.

국방혁신 4.0 과제별 중요도 평가 설문은 국방혁신 4.0 추진과 관련된 전문가 00명으로 선정하였으며, 국방혁신 4.0 전문성과 군 특수성을 고려하여 군 외부인사와 내부인사 비율을 0:0로 하여 전문가를 최종 선정하였다. 세부 인원수와

그룹 선정과 관련하여서는 보안 목적상 자세히 작성하지 않았다. 설문지는 총 00개 항목으로 구성하였으며, 평가 척도는 9점 척도를 활용하였다. 설문 전 연구팀의 Pilot Test를 통해 설문을 검증하고 이후 설문을 통해 관련 전문가에게 설문을 수렴하였으며 설문은 온·오프라인으로 실시하였다.



<그림 3-3> 국방혁신 4.0 중요도 평가를 위한 AHP 계층도

### 3.4 과제별 중요도 산출

전문가들로부터 종합된 설문을 이용하여 계층 내의 요소들의 쌍대비교를 통하여 계층별로 쌍대비교행렬을 구하였다. 아래 <표 3-1>은 ‘군 구조 및 교육 훈련 혁신’ 분야에서 각 과제의

상대적 중요도를 중요성 측면에서 설문한 예시이다. 본 논문에서 사용한 숫자는 군사보안 목적상 임의로 작성하였다.

<표 3-1>에서 쌍대비교행렬을 통하여  $\lambda_{max}$ 와 CI 값을 구해보면  $\lambda_{max} = 4.12$ ,  $CI = 0.04$ 로 일관성이 충족됨을 알 수 있다. 이렇게 각 전문가로부터 획득된 쌍대비교행렬을 계산하여 일관성을 검토하고 계층별 일관성이 충족되지 않은 (CI값이 0.1보다 큰) 설문 문항은 제외하였다.

이후 전문가 집단의 의견을 종합하여 가중치를 추정하기 위해 일관성을 충족하는 전문가들의 쌍대 비교행렬로부터 가중치를 구하여 이를 각각 기하평균으로 최종 가중치를 추정하는 수치 통합방법을 사용하였다. 이때 새로 구한 가중치의 합이 1이 되지 않으면 합이 1이 되도록 정규화하여 사용하였다. 이러한 방법을 거쳐 최종적인 가중치를 산출한 결과는 <표 3-2>와 같다. 보안상 과제명과 숫자는 임의로 작성하였다. <표 3-2>와 같이 계층별로 가중치가 추정되면 과제별 최종 가중치를 추정할 수 있게 된다. 예를 들어 <표 3-2>에서 ‘군 구조 및 교육 훈련 혁신’ 중점의 ‘첨단과학기술 기반 군 구조 발전’에 대한 ‘중요성’ 측면에서의 가중치를 구해보면  $0.5(\text{중요성 가중치}) \times 0.48(\text{중요성 측면에서의 군 구조 및 교육훈련 혁신 가중치}) \times 0.45(\text{첨단과학기술 기반 군 구조 발전 가중치}) = 0.108$ 가 되는 것이다.

<표 3-1> ‘군 구조 및 교육훈련 혁신’ 중점에 대한 설문결과

	첨단과학기술 기반 군 구조 발전	과학화 훈련체계 구축	예비전력 능력 확충	과학기술 인재 육성
첨단과학기술 기반 군 구조 발전	1	3	5	7
과학화 훈련체계 구축	1/3	1	3	5
예비전력 능력 확충	1/5	1/3	1	3
과학기술 인재 육성	1/7	1/5	1/3	1

<표 3-2> 계층별 최종 가중치 예

		전문가 1	...	전문가 n	최종 가중치
중요성		9.42	...	8.47	0.5
긴급성		4.39	...	1.26	0.3
효과성		1.96	...	3.24	0.2

		전문가 1	...	전문가 n	최종 가중치
중요성	북 핵·미사일 대응 능력 획기적 강화	9.13	...	6.81	0.07
	군사전략·작전 개념 선도적 발전	1.62	...	2.27	0.22
	AI 기반 핵심 첨단 전력 확보	3.91	...	6.81	0.19
	군 구조 및 교육훈련 혁신	3.18	...	4.72	0.48
	국방 R&D 전력 증강 체계 재설계	7.23	...	5.61	0.04

		전문가 1	...	전문가 n	최종 가중치
첨단과학기술 기반 군 구조 발전		6.07	...	4.11	0.45
과학화 훈련체계 구축		3.18	...	1.72	0.15
예비전력 능력 확충		1.56	...	1.47	0.05
과학기술 인재 육성		1.76	...	1.57	0.35

#### 4. 결론 및 향후 연구 방향

국방혁신 4.0이 추진되는 현시점에서 과제별 중요도를 평가하고 분석하는 것은 매우 중요하다. 본 연구에서는 AHP를 활용하여 프로젝트 소과제별 중요도를 평가하는 방법을 국방혁신 4.0 과제를 중심으로 제시하였다. AHP 방법은 다양한 현실적인 문제에도 적용되는 방법으로 평가 기준이 다수이고 복잡한 경우 전문가의 의견을 검증하고 평가요소의 중요도를 종합적으로 판단하는데 그 의미가 있다. 이번 연구에서는 AHP를 활용하여 과제별 중요도를 평가하는 방법만을 제시하였다. 향후에는 다음과 같은 연구를 후속할 예정이다.

첫째, 추진 중점 및 과제의 특성을 고려하여 정량적, 정성적, 정량 및 정성적 병행의 3개 성

과평가 유형으로 분류하여 평가하는 평가체계를 정립할 것이다. 둘째, 3가지 성과평가 유형으로 분류된 추진 중점 및 과제를 Bottom-up 방식으로 성과를 종합하여 중점·과제별 성과평가 결과를 본 연구에서 산출한 가중치를 활용하여 정량화된 진도율로 제시할 것이다. 셋째, 대형 프로젝트 성과관리 현황을 시각화하는 방법인 과학적 사업관리기법(EVMS)의 Bull's Eye Chart를 활용하여 프로젝트가 정상적으로 추진되고 있는지를 시각적으로 표현할 것이다.

향후 본 연구에서 시행하였던 과제별 중요도 평가 방법 적용하여 위와 같은 방식으로 국방혁신 4.0 성과평가를 방법을 개발한다면 객관적이고 종합적으로 프로젝트 진행 상황을 분석 평가하고 국방혁신 4.0의 추진현황을 직관적으로 이해하고 확인할 수 있을 것이다.

## 참고 문헌

- [1] 국방부, 『국방혁신 4.0 브로슈어』, 2023
- [2] 이상철, “국방개혁 2.0 진단과 발전방향”, 한국 군사문제연구원, 2021
- [3] 형혁규, “국방개혁 2.0의 평가와 향후과제”, 국회입법조사처, 2020
- [4] 최병욱, “국방개혁 추진, 어떻게 해야하나? : 탈냉전시대 미 육군의 개혁사례와 교훈”, 국방정책연구, 2019
- [5] 차두현, “국방개혁 2.0의 의미와 과제 : 설득력이 부족한 2.0으로의 변신”, 아산정책연구원, 2018
- [6] 김상범, “국방개혁의 추진에 따른 공중전력 발전과제와 방향”, 국방정책연구, 2006.
- [7] 차두현, “한국의 국방개혁 : 새로운 사고 새로운 체제를 통한 국방전환의 모색”, 제7회 항공력 국제 학술회의 논문집, 2004
- [8] 조근태, 조용곤, 강현수, 『계층분석적 의사결정』, 동현출판사, 2003
- [9] 이용복, “서열척도를 이용한 AHP 그룹의사결정 방법 - 무기체계 획득 과정을 중심으로”, 한양대학교 석사학위논문, 2006
- [10] 국방부, 『국방개혁 2020』, 2023
- [11] 국방부, 『국방개혁 기본계획 2009~2020』, 2009
- [12] 국방부, 『국방개혁 기본계획 2012~2030』, 2012
- [13] 국방부, 『국방개혁 2.0 강화군대 책임국방 구현』, 2018

## 저자 소개



**강승욱 (E-mail: seunguk0902@gmail.com)**

2015 육군사관학교 운영분석 학사  
현재 국방대학교 군사운영분석전공 석사과정  
관심분야 : 빅데이터 분석, 데이터과학,  
국방 M&S, 국방 AI



**윤진성 (E-mail: dbswlstjd92@naver.com)**

2015 육군사관학교 운영분석 학사  
현재 국방대학교 군사운영분석전공 석사과정  
관심분야 : 빅데이터 분석, 국방 M&S,  
인공지능, 데이터 과학



**소준호 (E-mail : so14246@naver.com)**

2003 육군사관학교 토목공학 학사  
2013 국방대학교 국제관계 석사  
현재 국방부 국방개혁실 군구조혁신분석담당  
관심분야 : 국방혁신 4.0, 군구조



**이용복(E-mail: miliman@naver.com)**

1997 육군사관학교 정보공학 학사

2006 한양대학교 산업공학 석사

2011 국방대학교 군사운영분석학 박사

현재 국방대학교 국방과학학과 교수

관심분야 : 비용분석, 시험평가, 국방 AI



설문조사 시 특이한 답변 결과에 대한  
원인 분석 사례 연구  
- 범국민 안보 의식조사 한미연합훈련의 필요성 답변 결과를  
중심으로 -

A Case study of analyzing the cause of unusual survey responses  
-Focusing on the necessity of the ROK-U.S. Joint Training  
in the national security consciousness survey-

운영빈1) · 문호석2)  
Yeongbin Yun · Hoseok Moon

ABSTRACT

The social perception of the necessity of the ROK-US joint training is an important factor in assessing the security consciousness of citizens. A unique finding was observed in a nationwide survey on security consciousness conducted by the Research Institute of National Security Affairs (RINSA) in 2022.

Various methodologies exist to analyze such social phenomena. This study employed statistical analysis and text mining techniques using defense-related domestic news articles, to interpret these phenomena. In conclusion, we observed a social bias phenomenon where individuals tend to conceal opposing opinions and respond with disinterest or neutrality, even if they do not agree with the ROK-US joint training.

Keywords : Security consciousness survey, Statistical testing, Text Mining,  
Sentiment Analysis, Analyze social phenomena, Social bias

---

논문접수일 : 2023년 5월 16일, 심사일 : 2023년 5월 31일, 게재확정일 : 2023년 6월 7일

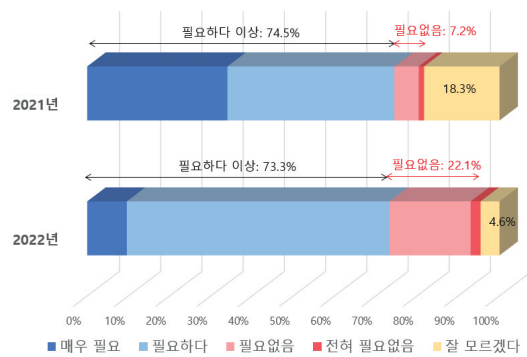
1) 국방대학교 국방과학학과 석사과정

2) 국방대학교 국방과학학과 교수 / 교신저자(hsmoon0329@gmail.com)

# 1. 서론

국방대학교 국가안보문제연구소에서는 매년 일반국민과 전문가의 안보의식을 종합적으로 파악하기 위해 ‘범국민 안보의식조사’를 실시하고 있으며 국민들의 안보의식 수준, 국가안보에 위협을 줄 수 있는 대외요인(강대국의 패권전쟁, 세계 경제 침체, 북한 군사적 위협 등), 국가안보 관련 일반인 및 군인 등 안보와 관련한 다양한 주제에 대한 설문조사 결과를 조사하여 분석한다[1].

본 연구에서는 2022년도 범국민 안보의식 조사에서 2021년도 결과와 특이한 차이를 보이는 일반국민 대상 ‘한·미 연합훈련의 필요성’에 대한 응답의 원인을 분석하였다.



<그림 1-1> 한·미 연합훈련의 필요성 (2021~2022)

<그림 1-1>은 우리나라 국민 성인남녀 1,200명을 대상으로 한·미연합훈련의 필요성에 대해 질문한 설문조사 결과<sup>3)</sup>이다. 2021년과 2022년 모두 한·미 연합훈련의 필요성에 대해 약 74%가 필요하다 이상의 응답을 하였다. 하지만 필요없음 이하 응답은 2022년도에 22.1%로 2021년 7.2%에 비해 3배이상 높았다.

특히 이같은 현상은 ‘잘 모르겠다’는 응답

의 연도별 차이, 즉 2021년도에 18.3%였던 ‘잘 모르겠다’가 2022년에 4.6%로 줄은 것과 관련이 있다. 왜 ‘잘 모르겠다’는 응답이 줄고, ‘필요없다’는 의견이 늘었을까? 이 질문에 대한 원인을 분석해 보고자 하는 것이 이번 연구의 목적이다.

본 연구에서는 이러한 원인을 분석하기 위해 표본의 대표성 및 설문조사 방식과 질문 차이에 대한 검정과 타 기관의 유사 질문에 대한 결과와의 비교, 그리고 일반적으로 나타나는 통계조사의 특징과의 관련성 등을 분석하는데 중점을 두었다.

또한 대량의 텍스트 데이터에서 통계적 분석으로 유의미한 정보를 추출하는 텍스트마이닝 기술을 활용하여 설문조사 결과와 국내 뉴스기사 데이터를 분석한 결과를 바탕으로 상호 연관성을 정량적으로 분석하여 사회현상을 해석해보고자 하였다.

본 논문은 크게 다섯 장으로 구성되었다. 2장에서는 범국민 안보 의식조사 및 한·미연합훈련의 역사와 의의를 소개하였으며 텍스트마이닝에 관한 개념과 기존 연구를 살펴보았다. 3장에서는 연구 절차와 방법을, 4장에서는 데이터 전처리 및 감성분석 과정 등 세부적인 결과를 시각적으로 표현하여 분석 결과를 제시하였으며, 설문조사에서 나타난 사회현상을 해석하였다. 마지막으로 5장에서는 유사한 결과를 보인 사례를 비교하여 한계점 및 추후 연구 방향에 대해 제시하였다.

## 2. 문헌 연구 및 관련 내용 정리

### 2.1 범국민 안보의식조사

국방대학교 안보문제연구소는 ‘범국민 안보의식조사 및 정책대안연구’ 중 일반국민과 군인의

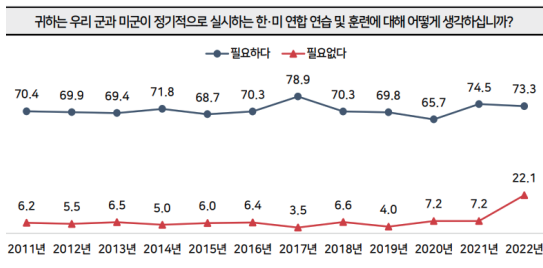
3) 만 18세~75세 성인남녀 1,200명(표본 오차: ±2.76%p) / 태블릿 PC를 이용한 면접조사

안보의식에 관한 조사를 실시하는데, 조사의 목적은 여론을 수렴하여 정부의 바람직한 안보정책 수립에 필요한 기초자료를 제공하는데 있다.

최현진 등(2019)은 30년간의 범국민 안보의식 조사 데이터베이스를 분석 대상으로 하여 조사 결과를 바탕으로 안보의식의 장·단기 변화 추세를 분석한 바 있는데, 이를 통해 여러 분야에 대한 시민 의식을 분석하여 안보의식의 변화요인을 파악하였다. 이러한 분석 결과는 범국민 안보의식 조사 결과가 국방정책 의사결정에 유용한 정보를 제공하고, 정책 대안을 제시하는데 활용될 수 있음을 시사하였다.

범국민 안보의식조사의 전체적인 구성은 국가안보 관련 일반사항, 한·미 동맹과 주한미군, 북한 관련 사항, 국방정책 및 국방현안, 병역제도 및 병역문화로 나누어져 있으며, 조사 결과를 통해 국민들이 인식하고 있는 국가안보사항, 국민들의 안보의식 수준 등을 분석한 결과를 제시한다.

그중에서도 본 연구에서 분석한 조사 결과는 ‘한·미 동맹과 주한미군’절의 ‘한·미연합훈련의 필요성’ 항목이며 2021년, 2022년 일반국민 응답결과를 바탕으로 하였다[2].



<그림 2-1> 한·미 연합훈련의 필요성 (2011~2022)

<그림 2-1>는 2011년부터 2021년까지 한·미연합훈련의 필요성에 대한 조사로 ‘필요하다’ 이상 응답은 평균 71% 이상 꾸준히 유지되고 있으며 ‘필요없다’ 이하 응답은 평균 7.1% 수준에 그쳐왔다. 하지만 2022년에 ‘필요없다’ 이하 응답이 급격하게 증가하였음을 알 수 있다.

## 2.2 한미연합훈련의 역사 및 의의

한미연합훈련은 정전협정이 체결된 후 유엔 군사령부가 주관이 되어 실시해왔고, 1978년 한미 연합사령부가 창설된 이후 한미 국방장관 간의 합의에 따라 설치된 한미군사위협회의 「전략지시 1호」에 근거하여 한미 연합사령부가 주관하여 실시되고 있다. 한미연합훈련의 목적은 한반도에서의 전쟁 억제 및 연합 전비태세 강화를 위해 연합 방위체제를 근간으로 통합된 작전계획을 시행하는 훈련으로 종류에는 훈련 성격에 따라 울지 프리덤 쉴드(UFS), 키리졸브(KR) 등으로 실시된다[4].

한미연합훈련은 한미동맹의 결속력을 강화하고, 북한의 도발을 억제하며, 한반도의 평화와 안보를 유지하는 훈련이자 실제로 계획된 작전을 시행하여 군의 전시 작전 능력을 검증하는 훈련이기도 하다. 하지만 매년 훈련을 시행할 때마다 북한은 비난과 협박을 하거나 한반도에 전쟁 위기를 조성한다며 격한 반응을 내놓고 있다. 작년 한 해만 보아도, 한미연합훈련이 주로 시행되는 6월과 9월, 핵 추진 항공모함을 포함한 한미연합훈련 등에 반발하며 SRBM(Short-range ballistic missile, 단거리탄도유도탄)과 KN-25(초대형방사포)를 여러 발 발사하였다.

언제 일어날지 모르는 전쟁에 대비를 해야 하는 존재 목적을 가진 군(軍)에게, 각종 훈련은 같고 닦아야 할 무기(武器)에 비유할 수 있다. 그중 한미연합훈련이라는 무기는 가장 강력한 무기 중 하나이며 단 한 번도 관리를 소홀히 하면 안 된다. 이렇듯 한미연합훈련의 중요성을 인지하는 것은 안보 인식 수준과도 직결되어 있다고 할 수 있다.

## 2.3 텍스트마이닝에 관한 연구

국방 분야에서 텍스트마이닝에 관련한 다양한 연구들이 이루어지고 있다. 윤승진 등(2015)

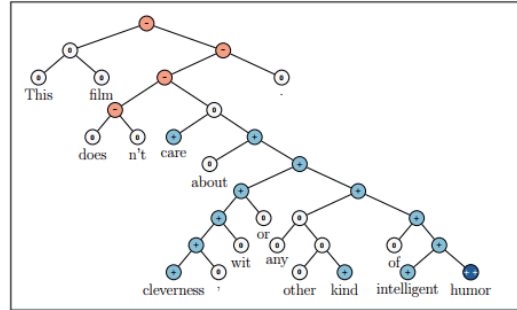
은 병사들의 생활지도기록부 자료와 SNS 텍스트 정보를 이용하여 토픽을 분석하고, 토픽 분석 결과를 독립변수로, 병사의 사고 유무를 종속변수로 두고 의사결정나무를 활용하여 사고 예측 모델을 개발하여 군 내 사고 예방 활동에 기여하였으며 이창용 등(2018)의 연구에서는 국내 뉴스 기사를 기반으로 북한의 핵실험에 대한 관심도를 분석하고 이를 토대로 토픽모형 분석을 통해 핵실험 관련 주요 토픽을 분석하였고 방한솔 등(2019)은 국내 뉴스 기사를 기반으로 북핵 관련 관심도를 분석하고 이를 토대로 토픽모형 분석을 실시하여 주요 토픽을 시계열적으로 표현하는 방법론을 제시하였다.

국방 분야 외에도 텍스트마이닝은 유용하게 사용되고 있다. 특히 최근에 주목받고 있는 OpenAI사에서 개발한 GPT(Generative Pre-trained Transformer) 기반 인공지능 챗봇 chatGPT는 수많은 텍스트 기반 데이터를 학습하여 인간의 언어를 학습하여 구축된 모델로 자연어 이해 및 생성에 대한 능력이 뛰어나 다양한 분야에서 활용되고 있다.

## 2.4 감성분석에 관한 연구

감성분석(Sentiment analysis)은 텍스트데이터가 내포하고 있는 긍정, 부정, 중립적인 감성을 컴퓨터로 분석하는 과정이며, 다양한 적용 분야와 방법론을 가지고 있다. 대표적으로 어떤 영화에 대해 사람들이 남긴 리뷰를 수집하여 영화에 대한 평점(0~10점)을 예측하여 영화에 대한 대중들의 평가를 보여주는 것, 마케팅 부서에서 SNS를 통해 상품에 대한 리뷰를 분석하여 새로운 전략을 수립하는 것 등이 대표적인 감성분석 사례라고 할 수 있다. 감성분석은 다양한 알고리즘을 이용하여 진행될 수 있는데, <그림 2-2>는 자연어 처리 분야에서 사용되는 신경망 모델 중 하나인 RNTN(Recursive Neural Tensor Network) 알고리즘으로 영화리뷰

텍스트를 토큰단위로 분해하여 분석하고, 다시 토큰을 합성하여 전체 문장에 대한 감성을 파악하는 순서로 최종적인 감성을 예측한다.



<그림 2-2> Recursive Neural Tensor Network<sup>4)</sup>

홍힝찬 등(2021)은 2010년부터 2019년까지 국방 관련 기사의 내용을 자연어처리 및 감성 사전을 이용하여 국방정책에 대한 사회적 인식을 분석하고 내부 제도 개선과 같은 사회적인 요구를 도출해내었으며, 허세영 등(2022)은 국내 뉴스 기사의 댓글에 대해 감성분석을 해서 비혼출산이라는 특정 주제에 대한 사회적인 인식을 분석하였고, 이에 대한 쟁점을 분석함으로써 현제도의 문제점이나 발전 방향 등을 제시하였다.

## 3. 연구 방법

### 3.1 연구 대상

본 연구는 2021년과 2022년의 범국민 안보의식 조사 중 「한·미연합훈련의 필요성」에 대한 일반국민의 응답결과를 기반으로 하였다. 이를 바탕으로 표본의 대표성과 설문조사 방식, 질문이 적절하였는지를 분석하였다. 그리고 아산정책연구원의 정례조사 및 한국국방연구원의 정례조사에서 실시한 ‘한국인의 한미관계 인식 조사’와 ‘국방에 대한 국민 의식 조사’ 결과 분석을

4) Data ScienceBlog Post (<https://blog.paralleldots.com/category/data-science>)

통해 비슷한 현상이 나타나는지 확인하였다.

또한 일반적으로 나타나는 통계조사의 특징과 연관성을 분석하였는데, 이를 위해 다양한 사회적 이슈에 대한 설문조사 결과와 사회적 감정을 비교하였다. 사회적인 감정은 텍스트마이닝을 활용하여 국내 뉴스기사 데이터로부터 해당 사회적 이슈에 대한 감성점수를 산출한 결과물이다.

먼저 한미연합훈련의 사회적 감정을 분석하기 위해 수집한 데이터는 네이버(NAVER) 뉴스에서 제공한 정치 카테고리 ‘국방/외교’ 분야의 기사이다. 네이버는 84개에 달하는 언론사로부터 뉴스를 제공받아 게시하고 있으며, 모든 정치적 성향의 매체를 포함하므로 우리나라에서 생산된 언론보도 내용 대부분을 포함하고 있다고 생각해도 무방하다고 생각된다. 따라서 2020년 1월~9월, 2021년 1월~9월, 2022년 1월~9월 간의 네이버 뉴스 국방/외교 분야 기사를 <표 3-1>과 같이 수집하였다.

<표 3-1> 국방/외교 분야 기사 데이터

date	title
20200101	트럼프 “김정은, 약속 지키는…”
20200101	‘정면돌파’ 23번 언급한 北…
20200101	신년사도, 남한 언급도 없는 김정은…
20200101	김정은 “새 전략무기 곧 선보여…”
20200101	라구람 라잔 시카고대 교수…
…	…

데이터는 192,732개의 행과 2개의 열로 구성되어 있으며 각 열은 날짜, 기사 제목을 의미한다. 전체 기사 중에 동영상 기사와 같이 제목이 없는 기사 및 중복된 기사는 제외(11.7%)하고 총 170,203건의 기사를 연구 대상으로 수집하였다. 또한 ‘한·미연합훈련’이라는 주제에 관한 기사만을 다시 분류하여 최종적으로 1,770개의 기사를 대상으로 하였다.

### 3.2 연구 절차와 방법



<그림 3-1> 연구절차

본 연구의 절차는 <그림 3-1>과 같다. 첫 번째로 조사대상, 표본수, 조사지역 등에 대해 문제가 없었는지 확인하기 위해 표본의 대표성을 통계적으로 검증하였다. 먼저 추출방법의 적절성을 판단하고, 여론 조사에서 중요한 표본의 지역 대표성을 2표본 T검정을 통해 검증하였다. 또한 2021년과 2022년도의 설문조사 방식이나 질문 및 답안 문항의 변화 여부의 유무도 확인하였다. 두 번째로 타 기관에서 동일연도에 실시한 유사 질문에 대한 조사결과를 비교분석하여 어떠한 유사점이나 차이가 있는지 살펴보았다. 마지막으로 일반적으로 나타나는 통계조사의 특징과 연관성을 분석하였는데, 텍스트마이닝을 이용하여 사회적 감정을 추출하여 비교함으로써 최종적으로 사회현상이 발생한 원인을 해석하였다.

## 4. 연구 결과

### 4.1 표본의 대표성 검증 및

#### 설문조사 방식 적절성 분석

차이를 유발할 수 있는 원인 중 하나는 조사 당시 표본의 대표성이 적절치 않거나 설문조사 방식 및 질문의 차이가 존재할 때이다. 확인 결과 표본은 지역별 인구비례 다단계 층화추출법으로 추출하였으며, <표 4-1>과 같이 지역 대표성 검증 결과 모집단에 비례하여 표본 크기는 잘 선정되었음을 볼 수

있다. 이는 두 집단의 2표본 T 검정을 통해 통계적으로 차이가 없음을 확인하였다.

<표 4-1> 지역 표본 대표성 확인 (단위: 명)

지역	모집단 비례 표본 크기	실제 표본 크기
계	1,200	1,200
서울	222	225
경기/인천	385	386
부산/울산/경상	180	179
대구/경북	116	121
광주/전라	117	114
대전/충청/세종	129	127
강원/제주	51	48

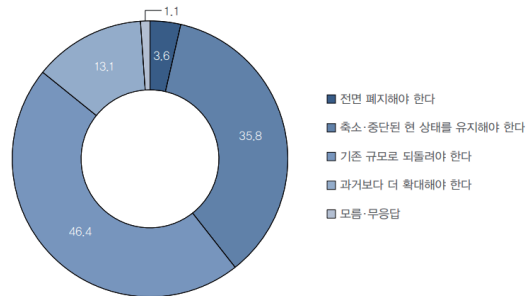
다음으로 2021년도와 2022년도의 설문조사 방식의 차이를 확인한 결과, 일반인 조사대상이 ‘만 18세~75세 성인남녀 1,200명(표본오차: ±2.76%p)’으로 동일하였고, 조사 방식 및 기간은 ‘테블릿 PC를 이용한 면접조사’와 ‘매년 9~10월’로 동일하여 조사대상과 표본 크기, 조사 방식에서는 차이가 없었다. 설문조사시 질문과 답안은 ‘귀하는 우리 군과 미군이 정기적으로 실시하는 한미연합연습 및 훈련에 대해 어떻게 생각하십니까?’와 ‘매우 필요-필요하다-잘 모르겠다-필요없다’로 동일한 질문 및 평정형(Rating) 답안이 사용되어 차이가 없었다.

## 4.2 타 기관의 유사 질문에 대한 조사결과 분석

타 기관의 유사 질문에 대한 조사결과가 범국민 안보의식 조사와 비슷한 결과가 관찰되는지 비교해 보았다. <그림 4-2>는 아산정책연구원의 정례조사<sup>5)</sup>인 ‘한국인의 한미관계 인

식 조사’에서 ‘한미 연합훈련 어떻게 해야 하나?’라는 질문에 대한 결과이다. ‘기존 규모 이상으로 해야한다’는 응답이 59.4%, ‘축소/중단된 상태 또는 전면 폐지해야 한다’는 응답이 39.4%였다.

비록 범국민 안보의식 조사와는 질문 문항에 차이가 있어 직접적인 비교는 제한되나, 마찬가지로 ‘잘 모르겠다’는 응답이 1.1%로 매우 낮은 것을 확인할 수 있었다[3].



<그림 4-1> 한·미 연합훈련에 대한 의견

한국국방연구원의 정례조사<sup>6)</sup>인 ‘국방에 대한 국민 의식 조사’에서는 2021년 질문 중 ‘한반도 전쟁 억제를 목적으로 한 정기적인 한미 연합훈련을 실시해야 하는가?’라는 질문에 대한 응답으로 ‘국방력 강화를 위해 계속 시행해야 한다’는 응답이 75.2%, ‘남북관계 개선을 위해 중단해야 한다’가 18.7%, ‘잘 모르겠다’가 4.1%였다.

질문 자체에 유도성 문구가 있음에도 불구하고, 한·미 연합훈련에 대한 부정적인 응답이 같은 해인 2021년도 범국민 안보의식조사(7.2%)보다 많이 높았다. 즉 범국민 안보의식 조사의 부정적인 응답이 다른 조사에 비해 낮은 편이라는 해석을 할 수 있다.

5) 2022년, 아산정책연구원에서 전화면접조사로 일반인 1,000명을 대상으로 실시

6) 2022년, 한국국방연구원에서 가구방문조사로 일반인 1,000명을 대상으로 실시

### 4.3 텍스트 데이터 전처리 및 탐색적 자료분석 결과

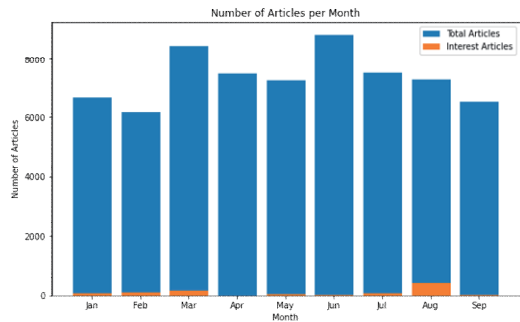
수집된 기사 제목은 한글, 한자, 숫자, 특수문자 등으로 구성되어 있는데 특수문자 및 한자, 숫자 등은 본 연구목적에 특별한 의미를 갖지 않는다고 판단하였기에 제거하고 한글 텍스트데이터만 보존하였다. 또한 완벽하게 명사로 분리되지 않은 단어들에 대해서는 조사(Postposition)를 제거하거나 단어를 보완하여 의미를 갖는 명사 형태로 보완하였고, 다양한 형태들로 표현되는 단어들을 하나의 형태로 변환시키는 스템밍(Stemming) 과정을 통해 분석에 정확도를 높이고자 하였다. 이러한 텍스트데이터의 형태소분석 및 전처리과정은 R프로그램의 KoNLP, tm, stringr 라이브러리를 이용하였다.

<표 4-2> 텍스트 데이터 전처리

기존(~에서)	변경(~으로)
한미연합훈련에	한미연합훈련
연합훈련시	한미연합훈련
미훈련	한미연합훈련
연합공중	한미연합공중훈련
연합지휘소	한미연합지휘소훈련
...	...
미훈련	한미연합훈련
연합훈련과	한미연합훈련

수집된 170,203개의 기사는 연도별로 53,598개(2020년), 59,898(2021년), 56,707(2022년)으로 한 해 평균 56,734개의 기사가 생산되었고, 3개년 모두 평균으로부터 1표준편차 내의 범위에 속하는 균등한 개수라고 할 수 있다. 이 중에서 한미연합훈련 관련 기사의 수는 각각 521, 963,

652개였는데, 연도별 전체 기사의 0.9%, 1.5%, 1.2%를 차지하였다. 이 중 2021년 8월에 다른 월에 비해 기사 수가 유난히 많았는데 2021년 7월 말부터 화두가 되었던 한미연합훈련 연기 및 축소 시행에 대한 주제와 8월에 김여정의 한미연합훈련 노골적 중단 요구, 훈련에 대한 유감 표명 논란<sup>7)</sup> 등 한미연합훈련에 관한 관심과 논란이 <그림 4-2>에서 보는 것과 같이 다른 달에 비해 높았기 때문으로 보인다.



<그림 4-2> 2021년 전체 기사/관심 기사 수 그래프

### 4.2 감성분석 결과

감성분석은 크게 머신러닝 기반 감성분석과 규칙 기반 감성분석으로 구분할 수 있는데, 전자의 경우에는 학습데이터에 감성 정보가 미리 레이블링 되어있어 이를 기반으로 새로운 텍스트데이터의 감성을 예측하는 지도학습 기반 감성분석이 대표적이다. 후자의 경우에는 감성 사전<sup>8)</sup>을 이용하여 정해진 규칙에 따라 텍스트데이터의 감성을 분석하는 방법이다. 전자의 경우 예측을 하고자 하는 데이터와 같은 도메인에서 생산되고 레이블링(Labeling)이 되어있는 다량의 데이터가 필요하다. 대표적으로 <표 4-3>에 나와있는 ‘Bag of Words Meets Bags of Popcorn<sup>9)</sup>’에서 제공한 IMDB<sup>10)</sup> 영화리뷰 데이터가 있다.

7) 연합뉴스 (<https://n.news.naver.com/mnews/article/001/0012570317?sid=100>)

8) 특정 단어에 대해 긍정, 부정, 중립으로 구분된 리스트

9) Kaggle의 데이터사이언스 경연대회 중 하나

10) Internet Movie Database의 약자. 영화, TV 프로그램 등과 관련된 정보를 제공하는 데이터베이스

<표 4-3>는 25,000개의 행과 3개의 열로 구성되어 있으며 각 열은 id(작성자 식별코드), senti(1: 긍정, 0: 부정으로 표현된 감성), review(실제 리뷰 텍스트)로 구분되어 있다. 따라서 목적에 맞는 알고리즘(로지스틱 회귀분석, 랜덤포레스트, 인공신경망 등)을 선택하여 각각의 리뷰는 긍/부정이라는 것을 학습시키면 새로운 영화리뷰에 대해서 긍정인지, 부정인지를 예측할 수 있게 되는데 이 방식이 머신러닝 기반 감성분석이다.

<표 4-3> IMDB 영화리뷰 데이터

id	senti	review
5814_8	1	With all stuff going down at the moment...
9495_8	1	Superbly trashy and wondrously unpretentious...
...	...	...
10095_3	0	Guy is a loser. Can't get girls, needs to build...
10194_3	0	This 30 minute documentary Buñuel made in the...

하지만 본 연구에서는 동일한 분야에 대해 레이블링된 학습데이터가 없었고 분석 결과에 대한 해석이 중요했기 때문에 감성 사전을 이용한 감성분석 방식을 선택하였다.

감성 사전기반은 감성 사전의 선택이 매우 중요하다. 어떤 도메인에서도 사용될 수 있는 보편성을 가져야 하기 때문이다. 따라서 감성 사전은 언어학과 관련된 전문가들이 형용사, 부사, 동사, 명사 등에 대한 단어를 다양한 방식으로 분석한 언어학적 연구를 기반으로 한다. 영어 감성 사전은 SentiWordNet, AFINN 등과 같이 일반적으로 많이 사용되는 사전이 존재하며, 한국어 감성 사전은 KNU 한국어 감성사전, KOSAC 등이 있다. 일반적으로 통용되는 사전은 없어 연구목적에 맞게 연구자가 사전을 선택하면 된다. 본 연구에서는 한국어 감성 사전 중에서도 가장 최근에 개발

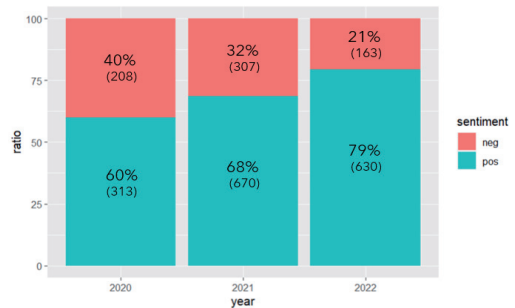
되었고 무료로 배포되어 누구나 활용할 수 있는 ‘KNU 한국어 감성 사전<sup>11)</sup>’을 이용하였다.

<표 4-4> KNU 한국어 감성 사전

word	polarity
가난뱅이	-2
가르침을 받아	0
가쁘히	1
...	...
설렁설렁	-1
설레게 하다	1
...	...

<표 4-4>과 같이 KNU 한국어 감성 사전 데이터는 14,855개의 행과 2개의 열로 구성되어 있으며 각각의 열은 단어나 문장을 의미하는 ‘word’와 감성점수인 ‘polarity’로 구성되어 있으며 긍정 또는 부정의 정도에 따라 점수가 2점에서 -2점까지 분포되어 있고 중립적인 단어의 경우에는 0점으로 평가되어 있다. 이러한 감성 사전을 바탕으로 관심 기사에 대해 긍정, 부정, 중립을 평가하였다.

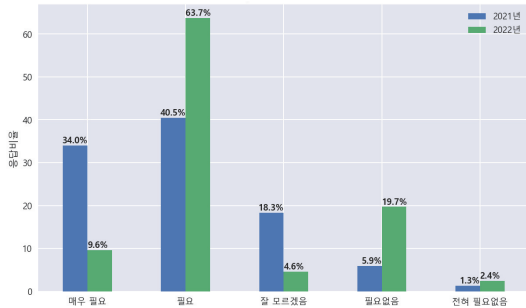
기사(Article)는 정의 자체에서 알 수 있듯 사실과 정보를 기반으로 작성되므로[5] 대부분의 기사 제목은 감정을 갖지 않는(감성 점수가 0점인) 중립 성격을 보였는데, 본 연구에서는 관심 기사에 대해 중립적인 제목은 혼련에 대한 관심도를 나타낸다는 측면에서 모두 긍정으로 분류하였다. 연도별 결과는 <그림 4-3>과 같다.



<그림 4-3> 2020~2022년 기사 감성분석 결과

11) 군산대학교 소프트웨어융합공학과에서 개발한 한국어 감성 사전

또한 2021년과 2022년 일반인 대상 설문 응답은 <그림 4-4>와 같은데, 설문조사 결과 ‘필요없음’ 이하 응답은 2021년도에 7.2%, 2022년도 22.1%로 14.9% 증가하였고 ‘잘 모르겠다’ 응답은 2021년도 18.3%, 2022년도 4.6%로 13.7% 감소하였다. 상대적으로 ‘잘 모르겠다’ 응답은 줄고, ‘필요없다’는 의견이 늘었다.



<그림 4-4> 2021, 2022년 설문조사 결과 비교

실제로 2021년에 비해 2022년에는 코로나19로 인해 축소되었던 한미연합훈련을 야외기동훈련(FTX)을 대폭 강화한 ‘을지 자유의 방패(UFS·Ulchi Freedom Shield) 연습’ 형태로 재개하는 등[6]의 국가안보를 위한 긍정적인 활동들이 많았음에도 오히려 부정적인 반응이 많았다.

통계조사의 특징 중 ‘사회적 바람직성(요망, desirability) 편향’이라는 현상이 있다. 이는 자신의 진짜 응답 대신에 사회적으로 바람직하고 정상적인 대답이라고 믿는 응답을 하거나, 자신의 긍정적 이미지를 나타내는 응답을 하는 것을 의미하는데, 선거일이 다가올 때 자기 의사를 솔직하게 표현하는 현상이 줄어드는 사회적 바람직성 편향 현상이 발생하는 것이 대표적이다. 이는 응답에 관한 내용이 사회적으로 민감한 주제일 때 더 심하게 나타난다. 왜냐하면 자기 생각이 사회적으로 용납하기 어려운 경우 다른 사람으로부터 비난이나 반발을 받을 수 있기 때문이다.

또한, 2021년도와 2022년도 1~9월간 한미 연합훈련 관련 뉴스 기사 감성분석 결과, 긍정 뉴

스 비율이 증대되어 2022년도에는 한미 연합훈련에 대해서 더 많이 강조되었다고 볼 수 있다. 그리고 동일 기간에 북한은 탄도미사일을 2021년에 3회, 2022년에 21회 도발하여 남북 관계 긴장이 증대되었다. 이러한 사실들을 근거로 할 때, 정부 정책과 남북 관계 긴장 증대에 대한 우려로 설문조사 응답에서 ‘모르겠다’ 응답이 줄어들 가능성이 내재되어 있다고 해석할 수 있다.

## 5. 결론

본 연구에서는 설문조사 결과에서 나타나는 사회현상을 해석하기 위해 통계적인 분석을 통해 검정하고, 텍스트마이닝의 한 분야인 감성분석을 이용하는 방법론을 제시하였다. 그 결과 특정 사회현상의 원인을 데이터를 다각도로 분석하여 결론을 도출해낼 수 있었다.

한미연합훈련이라는 주제 외에도 다른 사회현상에서 비슷한 결과를 보인 사례가 있다. 첫째는 전국장애인차별철폐연대(이하 전장연)의 사례이다. 전장연은 2007년 9월 출범한 단체로 출범 이후 장애인 인권 신장을 목표로 다양한 형태의 활동을 해왔다[7]. 그리고 대중의 관심을 받기 시작한 계기는 2021년 ‘지하철 탑승 시위’를 시작하면서부터이다. 이 시점을 기준으로 전/후 장애인의 인식에 관한 설문조사 결과를 보면 긍정 반응이 4% 감소, 중립이 5.5% 감소, 부정은 9.3% 증가하는 현상을 보였다[8][9].

두번째 사례는 원자력발전소 유지에 관한 사례이다. 원자력발전소(이하 원전)는 핵분열을 이용해 발전하는 발전소를 말하며, 다른 발전소에 비해 많은 양의 전력을 안정적으로 공급할 수 있는 장점이 있지만, 핵분열 과정에서 방사성 물질이 발생하므로 사고가 나면 막대한 피해가 발생할 수 있어 여러 가지 안전대책이 요구된다[10]. 하지만 2011년 3월 후쿠시마 원전

사고 이래로 원전에 대한 우려가 커지기 시작했고, 신고리 원전 반대 시위 등으로 표출되었다. 19대 대선에서는 공약 정책으로서 등장하기도 했다. 그 결과 2017년 6월, 국내 최초 원전인 고리원전 1호기를 완공 40년 만에 영구 정지되었으나, 현 정부는 반대로 탈원전 정책을 폐기를 내세웠고 가동을 멈추기로 했던 원전을 계속 가동하기로 수정하여 현재까지 진행되고 있다. 따라서 2017년과 2022년의 설문조사 결과를 비교해보았는데 긍정 반응은 비슷하였으나, 중립 의견이 10% 감소하였고 부정 반응이 8% 증가하였다[11][12].

하지만 사회현상은 매우 복잡하고 다양한 요소들이 상호작용하여 만들어지는 현상이다. 따라서 본 연구의 결과가 모든 사회현상에 적용될 수 있다는 것을 보장하지는 않는다. 즉, 단일한 방법으로 해석하면 그 현상의 복잡성을 충분히 반영하지 못할 가능성이 있으므로 다양한 시각과 접근 방법을 활용하여 종합적으로 판단하는 것이 필요하다.

따라서 본 연구도 특정 사회현상을 해석하는 방법론 중 하나로서 특정 정책에 대한 의사결정을 지원하거나 여론의 반응을 사전에 예측하여 국민의 요구를 만족하는 정책을 시행하는데 기여할 수 있을 것으로 기대된다.

## 참 고 문 헌

- [1] 국방대학교 국가안전보장문제연구소, 『범국민 안보의식 조사개요』, 2022
- [2] 국방대학교 국가안전보장문제연구소, 『범국민 안보의식 조사결과 요약』, 2022
- [3] J. James Kim, 강충구, 함건희, “한국인의 한미관계 인식”, 2022
- [4] 양영조, “한미연합연습”, 군사편찬연구소, 2015.
- [5] 국립국어원, 『표준국어대사전』, 2023
- [6] 국방일보, 『한미 ‘을지 자유의 방패 연습’ 22일 시작』, 2022
- [7] 전국장애인차별철폐연대, 『전장연이 자나온길』, 2023
- [8] 서울신문, 『장애인 인권에 대한 인식조사 결과』, 2019
- [9] 미디어토마토, 『선거 및 사회현안 32차 정기 여론조사 결과』, 2022
- [10] 한전원자력연료, 『월자력발전소 개념도』, 2022
- [11] 천지신문 『탈원전 정책에 대한 찬성·반대조사 결과』, 2017
- [12] 전자신문, 『에너지 정책 방향 국민 여론조사 결과』, 2022

## 저 자 소 개



**윤영빈(E-mail: yb6470@gmail.com)**

2014 육군사관학교 토목환경공학 학사  
현재 국방대학교 군사운영분석전공 석사과정  
관심분야 : 텍스트 마이닝, 데이터과학, 국방 AI



**문호석(E-mail: hsmoon0329@kndu.ac.kr)**

1994 육군사관학교 화학과 학사  
2003 고려대학교 전자공학 석사  
2006 고려대학교 산업공학 박사  
2010 고려대학교 통계학 박사  
현재 국방대학교 국방과학학과 교수  
국가안전보장문제연구소  
군사과학연구센터장  
관심분야 : KCTC 빅데이터분석, 데이터과학,  
국방 AI, M&S



한 글 제 목(굴림 16)

영 문 제 목(신명조 12)

이센터1) · 김센터2)(굴림 11)

Cen-Ter Lee · Cen-Ter Kim (신명조 11)

**ABSTRACT(견명조 10)**

abstract abstract abstract abstract abstract abstract(신명조 10)

Keywords : Keywords, Keywords, Keywords, Keywords, Keywords, Keywords,  
Keywords,

---

1) 00대학교 0000전공 석사과정(바탕 9)  
2) 00대학교 0000전공 교수

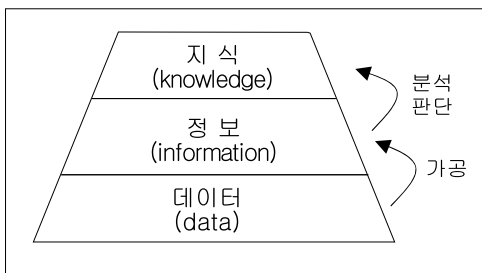
### 1. 서 론(HY중고딕 14)

현대사회에서 국방기술은 ~~~~~~  
 ~~~~~~ 연구방안 수립과 추진이 필요하다.  
 국내외의 ~~~~~~필요성의 증대로  
 귀결된다.  
 이처럼 ~~~~~~  
 ~~~~~~ 하고 있다.  
 또한 과학기술의 ~~~~~~  
 ~ Scientometrics'라 한다.[8] ~~~~~~  
 ~연구를 의미한다.(신명조 10)

### 2. 000 고찰

#### 2.1 0000000(휴먼고딕 13)

과학기술 연구활동의 ~~~~~~  
 ~~ 정보이다.  
 이런 ~~~~~~  
 ~~~~~~부분이다.[7]  
 그러므로, ~~~~~~필요  
 하다. <그림 1>은 ~~~~~~  
 보여준다.



<그림 1> 데이터, 정보, 지식의 계층 구조

데이터는 ~~~~~~  
 ~~~~~~올라가게 된다.[6]

#### 2.2 0000

최근의 ~~~~~~  
 ~~~~과정이라 할 수 있다.

최근 ~~~~~~  
 ~~~~개괄적인 비교는 <표 1>과 같다.

<표 1> 000000 비교

|  |  |  |
|--|--|--|
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

### 3. 00000

현재 ~~~~~~  
 ~~~~~~ 있다.

### 4. 000 방법

연구동향을 ~~~~~~  
 ~~~~~~같다.

#### 4.1 00000

1차 ~~~~~~  
 ~~~~~~있다.

#### 4.2 00000

1차 ~~~~~~  
 ~~~~~~수집하였다.

### 5. 00결과

#### 5.1 0000000

자율주행 ~~~~~~  
 ~~~~사용하였다.

#### 5.2 00000

지형/물체 ~~~~~~  
 ~~~~입력하였다.

### 5.3 00000

상위 ~~~~각주3)~~~~~  
~~~~~ 있다.

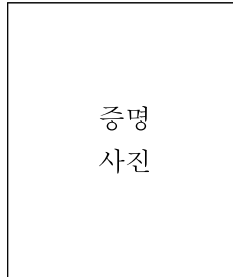
## 6. 결 론

지금까지 ~~~~~~  
~~~~~기대된다.  
그러나 ~~~~~~  
~~~~~필요하다.

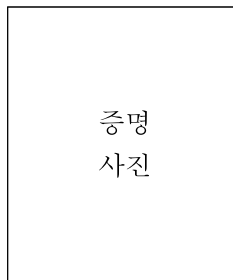
## 참 고 문 헌(휴먼고딕 16)

- [1] 국방기술품질원, 『2007 국방과학기술조사서(일반본) 제4권』, 국방기술품질원, 2012
- [2] 이주장·김현진·이민철·강정원·권인소·송재복, “차울주행기술”, 기계저널 제47권, 2007
- [3] 박용태, 『기술과 경영』, 생능출판사, 2005
- [4] 임치환, “Knowledge Map의 활용에 관한 연구”, 한국외국어 대학교 석사학위논문, 2006
- [5] 윤문섭·이우형·김윤명·오해영·손성혁, “친기술 연구기획 사전 타당성 분석을 위한 지식맵 작성 방법론 개발 및 활용방안”, STEPI, 2003

## 저 자 소 개(HY신명조 13)



000(E - mail: 0000000@naver.com)  
 2000 0000000 졸업(문학사)  
 현재 00대학교 0000전공 석사과정  
 관심분야 : 데이터마이닝, OR/SA,



000(E - mail: 00000000000@0000000)  
 1991 00000학교 졸업(이학사)  
 1997 미국 UC. Berkeley 졸업(0000 석사)  
 2005 KDI 00대학원 졸업(000000 석사)  
 2006 00대학교 졸업(0000 박사)  
 관심분야 :

---

3) 각주내용.



| 발행인 |

김홍석(국방대학교 총장)

| 편집인 |

박영준(국가안전보장문제연구소장)

---

## 군사과학연구

제16권 제1호

---

2023년 6월 30일 인쇄

2023년 6월 30일 발행

발행처 : 국방대학교 국가안전보장문제연구소

TEL. (041) 831-6414

E-mail. rinsakj@kndu.ac.kr / kndu212@kndu.ac.kr

인 쇄 : 청 맥 기 획 (042) 487-2589

---

ISSN 1975-3888



## Research Papers

Case Study Approach to Data Wrangling Techniques for Modeling and Simulation  
/ **Bongkyoo Yoon**

A Study on Predicting the Readiness Posture Environment of North Korea's Ballistic Missile using Weather Information  
/ **Sungeel Jo · Choonjoo Lee · Hoseok Moon**

Method of evaluating the importance of tasks in a project using AHP (Analytic Hierarchy Process)  
/ **Seunguk Kang · Jinseong Yun · Junho So · Yongbok Lee**

A Case study of analyzing the cause of unusual survey responses  
- Focusing on the necessity of the ROK-U.S. Joint Training in the national security consciousness survey -  
/ **Yeongbin Yun · Hoseok Moon**

