

# 의사결정나무를 이용한 인천국제공항에서의 안개발생 예측기법 개발

## The Prediction of Fog Occurrence at the Incheon Int'l Airport Using Decision Tree Model

임헌호\*, 이화운, 김동혁(부산대학교)

### 1. 서론

일반적으로 기상학에서의 안개는 지표면 근처에서 수증기 공급 또는 지표면 냉각으로 인해 수증기가 수적으로 응결되어 나타나는 현상으로 항공기상 관측 및 Oke(1987)는 수증기 응결로 인한 수평시정인 1,000m미만인 시정장애 현상으로서 안개를 정의하고 있다. 이러한 안개는 가시거리를 악화시킴으로 인해 인간 활동 가운데 교통 및 수송과 관련하여 큰 위험성을 내재하여 해안지역에서는 해무로 인한 선박 및 비행기 운항에 큰 영향을 미치고 있다. 특히, 많은 공항이 연안에 위치한 우리나라를 포함하여 중위도 지역의 많은 나라들에 있어 안개가 매년 많은 부분의 항공기 회항 및 연착을 초래하는 사실과 함께(Mason, 1982), 원할한 공항 운영 및 안전한 항공기 운항 측면에서의 항공기상학적 안개 연구 및 기상관측의 중요성을 밝힌 바 있다.

그러나 항공기 운항에로의 실시간 항공기상 지원이라는 궁극적인 목적과 활용성으로 인해 위의 순수 기상학에서의 연구들과는 다른 방법으로 연구가 수행되었다. 즉, 항공기상에 있어 안개에 관한 대부분의 연구는 공항에서 발생한 안개의 기후학적 특성과 함께 안개 발생과 소산에 대한 다른 기상요소들 특징을 토대로 통계적인 접근으로 예측하고 지원하였다.

따라서 본 연구는 인천국제공항을 대상으로 항공기상 지원에 있어서 가장 중요한 안개발생 예측기법을 개발하였다. 즉, 실시간 예측 및 지원을 위해 관측자료를 토대로 의사결정나무를 이용한 실시간 예측기법을 개발하였다.

### 2. 연구방법 및 자료

본 연구에서는 공항이라는 특수성 및 활용성

으로 인해 인천국제공항에서 안개발생에 대한 실시간 예측 및 지원을 목적으로 통계모형 가운데 하나인 의사결정나무를 이용한 예측기법을 개발하였다. 즉, Breimann 외(1984)에 의해 제안되어 의학 분야를 비롯하여 많은 의사결정 문제 뿐 아니라 강수량 등과 같은 연속 변수량을 예측하는 문제에도 이용되는 의사결정나무 분석을 통해 인천국제공항에서의 안개발생에 대해 실시간 예측을 하였다.

Data Mining 응용프로그램에서 지원하는 모형 가운데 인천국제공항에서의 안개발생 예측이라는 연구목적과 성격에 맞게 먼저, 시간별 관측자료를 이용하여 우시정을 직접 예측하는 결정나무모형(Decision Tree for Predicting Prevailing Visibility, 이하 DT\_Pvis)을 분석(Training)하여 그 결과의 검증(Validation) 및 정확성을 통해 이용가능성을 살펴보았다. 그리고 일별 관측자료 및 시간별 관측자료를 이용하여 안개발생 여부를 예측하는 결정나무모형을 분석하였고 나아가 안개발생 메커니즘에 따라 그 모형을 세분화하여 살펴보았다. 즉, 안개발생 여부를 예측하는 결정나무모형은 먼저, 일별 관측자료를 이용하여 안개발생일에 대한 결정나무모형(Decision Tree of Foggy Day, 이하 DT\_F)을 분석하고 이 모형에 의해 안개발생일로 예측된 날에 대해서는 안개발생 메커니즘에 의거하여 이류무 발생일과 증기무 발생일로 구분하였다. 각각 구분된 안개발생일에 대하여 시간별 관측자료를 입력자료로 한 결정나무모형(Decision Tree of Advection Foggy Day, 이하 DT\_advF; Decision Tree of Steam Foggy Day, 이하 DT\_stmF)을 각각 분석하여 인천국제공항에서의 실시간 안개발생에 대한 예측기법을 개발하였다. 동시에 안개발생 메커니즘별로 구분하지 않고 동일한 시간별 입력자료를 이용한 결정나무모형(Decision Tree of Total Foggy Day, 이하 DT\_totF)과의 정확성을 비교

하였다.

그리고 이러한 각각의 결정나무모형을 분석함에 있어 공통된 과정은 다음과 같다.

(1) 자료추출 과정(Sampling Node)

관측자료들을 모형이 분석할 수 있는 형태인 Data Set으로 구성하는 과정으로 먼저, Data Set을 구성하는 변수들에 대해 입력변수(input variable)와 목표변수(target variable)로 그 역할을 지정하였다. 우시정을 예측하기 위한 모형(DT\_Pvis)에서는 연속형 변수인 우시정이 목표변수가 되며 안개발생 여부를 예측하기 위한 모형(DT\_F, DT\_totF, DT\_advF, DT\_stmF)에서는 이진형 변수인 발생여부가 목표변수가 된다. 그리고 각 Data Set을 분석용(Traning Data Set)과 검증용(Validation Data Set)으로 분할하되 임의발생함수(Random Seed)를 이용하여 각각 70%와 30%의 비율로 각 Data Set을 구성하였다.

(2) 탐색 과정(Exploration Node)

여러 측면에서의 데이터 탐색을 통해 기본적인 정보를 검색하고 유용한 정보를 찾아내는 과정으로 본 연구에서는 상관계수를 통해 목표변수와 관련이 적은 변수들을 미리 제거하였다. 즉, 연속형 변수인 우시정을 예측하기 위한 모형(DT\_Pvis)에서는 목표변수와의 R-Square 상관계수가 0.005 미만인 변수를 탐색 과정에서 미리 제거하였고 이진형 변수인 안개발생 여부를 예측하기 위한 모형(DT\_F, DT\_totF, DT\_advF, DT\_stmF)에서는 목표변수와 Chi-Square 상관계수가 0.005 미만인 변수를 탐색 과정에서 미리 제거하여 모형 분석 및 검증에 있어 효율성을 높였다.

(3) 모형화 과정(Modeling Node)

본 연구에서는 이진형(Binary), 명목형(Nominal), 순서형(Ordinal), 연속형(Interval) 변수들에 대해서 수행 가능한 결정나무 알고리즘을 토대로 모형을 분석하였다. 즉, 연속형 변수인 우시정을 예측하기 위한 모형(DT\_Pvis)에서는 분리기준으로 F-Square 통계량을, 이진형 변수인 안개발생 여부를 예측하기 위한 모형(DT\_F, DT\_totF, DT\_advF, DT\_stmF)에서는 Chi-Square 통계량을 사용하여 모형을 분석하였다.

(4) 평가 과정(Assessment Node)

마지막으로 위와 같은 과정을 통해 분석된 통계모형은 좋은 예보모형이 되기 위하여 다음과 같은 두 가지 조건을 만족하여야 한다(손건태 외, 2005).

[조건 1] 모형 분석자료로 학습된 모형에 의한 검증에서 모형 분석결과와 모형 검증결과가 비슷해야 자료집단에 따라 둔감한(robust) 좋은 예측모형이 된다.

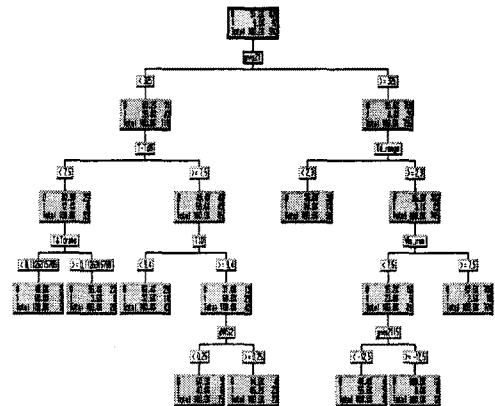
[조건 2] 기상예보에 활용될 수 있는 정확도와 설명력을 지녀야 한다.

즉, 분석 모형에 대한 신뢰성, 타당성, 유용성을 평가하기 위해 본 연구에서는 먼저, 모형분석결과와 모형검증결과를 비교하였고 또한, ROC 곡선과 Heidke Skill Score, 그리고 Treat Score를 이용해 모형의 정확성을 평가하였다.

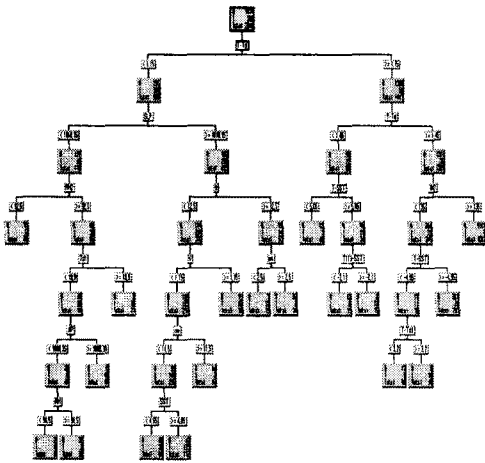
3. 연구결과

먼저 Fig. 1은 최근 4년(2000년 9월 1일 ~ 2004년 8월 31일) 동안의 비강수일 913일에 대한 안개발생일을 예측한 의사결정나무모형이다. 그리고 Fig. 2는 안개발생일 예측모형에 의해 예측된 안개발생일 85일에 대한 시간별 안개발생을 예측한 의사결정나무모형이다. 그리고 이에 대해서 안개의 발생메커니즘에 따라 이류무 및 증기무로 나누어 각각 구축한 의사결정나무모형은 Fig. 3과 Fig. 4에 나타내었다. 결과에서 처럼 노점편차와 U 및 V, 그리고 SST 등이 안개발생 여부를 결정하는데 매우 중요한 변수로 작용하고 있다.

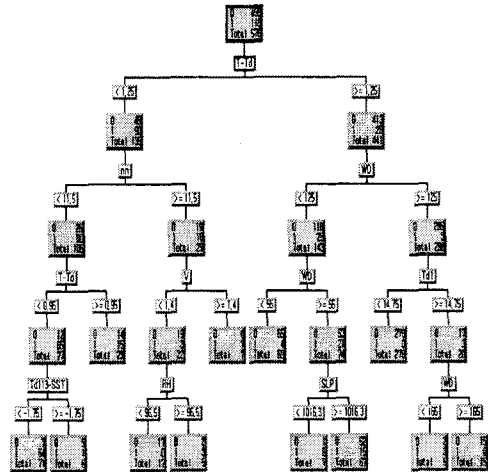
그리고 각각의 예측모형에 대한 정확성을 Heidke Skill Score와 Treat Score로 Table 1에 나타내었다. 특히, 안개발생일을 그 발생 메커니즘별로 구분하여 각각에 대한 시간별 안개발생 예측을 한 경우가 상당히 높아졌다.



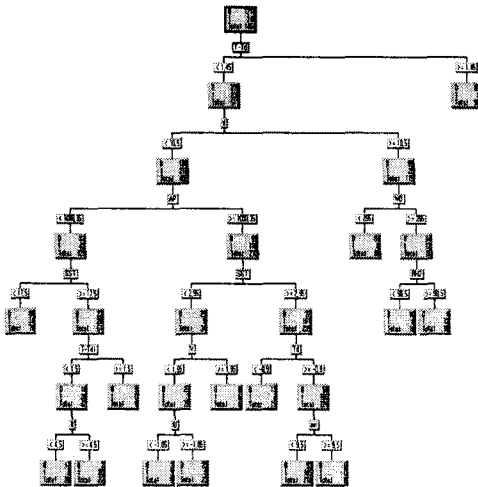
<Figure 1> Decision Tree for predicting the foggy day



<Figure 2> Decision Tree for predicting the foggy time without classifying fog type



<Figure 4> Decision Tree for predicting the foggy time with classifying steam fog



<Figure 3> Decision Tree for predicting the foggy time with classifying advection fog

Table 1. The Scores for predicting model

	Threshold Value	Heidke Skill Score	Treat Score
Foggy day	$T > 0.25$	0.62	0.51
All Fog(85)	$T > 0.4$	0.67	0.59
Advection Fog(61)	$T > 0.3$	0.76	0.67
Steam Fog(24)	$T > 0.3$	0.76	0.68

#### 4. 감사

본 연구는 한국학술진흥재단의 2004년도 신진연구인력지원사업(C00037)의 지원을 받아 수행되었으므로 이에 감사를 드립니다.

**참고문헌**

Breimann, L.,J. H. Friedman, R. A. Olshen and C.J. Stone, 1994: Classification and Regression Trees, Chapman & Hall, New York

Mason, 1982: The physics of radiation fog. J. Meteor. Soc. Japan, 60(1), 486-498.

Oke, T. R., 1987: Boundary Layer Climates. Methuen, London and New York, 240-241.