

# ADF를 사용한 유전프로그래밍 기반 비선형 회귀분석 기법 개선 및 풍속 예보 보정 응용

## Improvement of Genetic Programming Based Nonlinear Regression Using ADF and Application for Prediction MOS of Wind Speed

오 승 철\* · 서 기 성\*  
(Seungchul Oh · Kisung Seo)

**Abstract** - A linear regression is widely used for prediction problem, but it is hard to manage an irregular nature of nonlinear system. Although nonlinear regression methods have been adopted, most of them are only fit to low and limited structure problem with small number of independent variables. However, real-world problem, such as weather prediction required complex nonlinear regression with large number of variables. GP(Genetic Programming) based evolutionary nonlinear regression method is an efficient approach to attach the challenging problem. This paper introduces the improvement of an 해 based nonlinear regression method using ADF(Automatically Defined Function). It is believed ADFs allow the evolution of modular solutions and, consequently, improve the performance of the GP technique. The suggested ADF based GP nonlinear regression methods are compared with UM, MLR, and previous GP method for 3 days prediction of wind speed using MOS(Model Output Statistics) for partial South Korean regions. The UM and KLAPS data of 2007-2009, 2011-2013 years are used for experimentation.

**Key Words** : Genetic programming, Automatically defined function, Wind speed prediction, MOS(Model output statistics)

### 1. 서 론

개체를 스트링으로 표현하는 유전알고리즘(GA, Genetic Algorithm)[1]과는 달리, 유전프로그래밍(GP, Genetic Programming)[2-4]은 트리자료를 이용한다. 이에 따라 GP는 개체의 표현에 가변성을 가지게 되었고, 구조적으로 열린 해의 탐색에 적합하다. 이를 통해, 복잡하고 실용적인 설계와 최적화 문제에 많은 응용이 이루어지고 있다[4, 5].

그러나, 트리의 열린 구조로 인해 탐색공간이 방대해짐에 따라, 수행시 GA보다 탐색의 어려움을 겪고 있다. 개체의 표현에 트리구조를 이용하게 됨에 따라 bloat 코드[6], 파괴적인 교배방식[7], 구조적 어려움[8, 9] 등의 문제가 제시되었다. 이에 따라, GP의 탐색 성능을 향상시키기 위해서, 해의 표현, 선택, 유전연산 등에 대해서 연구가 진행되어 왔고, 문제에 따라서 각 부분의 영향이 서로 다른 것으로 알려지고 있다[3, 6-8].

본 논문에서는 GP의 대표적 응용분야인 기호적(symbolic) 회

귀분석을 다량의 독립변수(인수)를 가진 풍속 예측 보정에 응용하고자 한다. 해당 문제가 고도의 비선형 특성을 가지고 있으므로, 다중선형회귀분석보다 GP의 기호적 회귀분석을 이용한 열린 비선형 회귀분석의 접근이 적합하다. 이러한 접근법은 고정된 구조에 대해서 계수만 최적화하는 다중선형회귀분석보다 이상적이나, 회귀식의 구조와 계수를 동시에 최적화해야하기 때문에, 탐색 공간이 매우 방대하다는 단점이 있다. GP의 탐색 성능을 향상시키는 연구가 대부분 벤치마크 문제에 적용한 경우가 대부분이라, 실용적인 문제에 대해서 문제의 특성을 고려한 접근은 많지 않다.

본 논문에서는 다량의 인자를 포함하는 GP 기반 기호적 회귀 분석 문제에 대해서 해를 구성하는 함수와 터미널의 선택 문제에 대해서 고찰한다. 그리고, GP의 탐색성능을 개선하기 위해서, 자동정의함수(Automatically Defined Function, ADF)[3]라는 모듈화 개념을 적용한다. 인자들의 비선형 결합시에 공통적으로 나타나는 부분적인 수식 형태에 대한 모듈화를 통해 반복 사용이 가능하며, 같은 크기의 트리에서 더 복잡화된 연산 수행이 가능하다.

사례 연구로서 진화적 기호회귀 분석기법을 풍속 예보의 보정식 생성에 적용한다. 남한지역에 대해서 2007년~2009, 2011~2013년의 방대한 기상 데이터를 대상으로 다양한 최적화 실험을 수행하고, 풍속 예측에 대한 성능을 다중선형회귀분석법과 비교한다.

† Corresponding Author : Dept. of Electronics Engineering,  
Seokyeong University, Korea

E-mail : ksseo@skuniv.ac.kr

\* Dept. of Electronics Engineering, Seokyeong University, Korea  
Received : September 29 2015; Accepted : November 7, 2015

## 2. 비선형회귀분석과 유전프로그래밍

### 2.1 비선형회귀분석

선형회귀분석에 의한 접근으로는 기상을 포함한 자연계에 존재하는 수 많은 비선형 특성을 가진 시스템을 정확하게 모델링하기가 어렵다. 그림 1은 주어진 임의의 데이터에 대한 선형 및 비선형 회귀분석의 차이를 잘 나타내주고 있다.

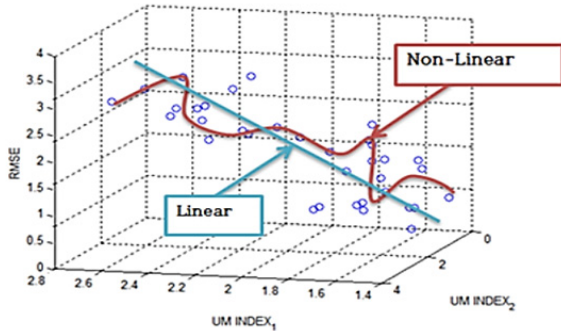


그림 1 선형회귀분석과 비선형회귀분석의 예  
Fig. 1 Example of linear and non-linear regression

그러나, 일반적인 비선형회귀 분석은 차수가 2차나, 3차와 같은 저차 다항식으로 고정되어 있으며, 정해진 차수 이내에서 계수 추정을 통해 오차를 최소화한다. 따라서, 그림 1과 같은 임의의 비선형성을 나타내는 데는 한계가 있으며, 이를 위해서는 차수 및 함수가 제한되지 않는 열려진 구조를 가진 비선형회귀분석의 접근이 필요하다.

### 2.2 유전프로그래밍

본 연구에 사용된 최적화 알고리즘은 유전프로그래밍(GP, Genetic Programming)[2, 3]으로서, 유전알고리즘(GA, Genetic Algorithm)과 달리 개체를 트리 형태로 표현하는 특징을 가지고, 유전자 크기가 가변이며, 구조적으로 열려진 공간의 탐색 문제에 적합하다.

GP의 상기 특성을 이용하면 구성요소가 제한되지 않으며, 복잡한 비선형 회귀분석에 적합한 다항식을 생성할 수 있다. 즉, 주어진 입력 쌍을 가장 잘 따라가는 다항식을 GP의 유전자 형태인 트리로 구성하기 때문에 다항식의 차수가 열려있고, 더욱이 사칙연산외에 초월함수 및 조건 연산자까지 구성요소로 선택할 수 있어서 제한이 없는 회귀분석식을 생성할 수 있다. 이를 기호적(symbolic) 회귀분석이라고 하며, 구조와 계수가 열려진 임의의 비선형회귀분석식의 생성이 가능하다.

GP의 수행과정을 유사 코드 형태로 나타내면 그림 2와 같다. 사전에 유전자를(트리) 구성하는 함수와 터미널을 정의한다. 함수와 터미널은 해를 조합적으로 구현할 수 있는 요소들을 선택한

다. 초기에 트리로 구성된 각 개체들을 임의로 생성한다. 그리고, 각 개체 트리를 해석하여 구한 후보 해를 적합도 함수로 평가한다. 이후, 주어진 선택 방법에 의해 유전연산에 참여할 개체들을 선택한다. 선택된 개체들을 대상으로 유전연산(교배, 돌연변이) 수행한다. 그리고 이 전체과정을 종료조건이 만족될 때 까지 반복한다. GP에서의 교배와 변이 연산자는 아래 그림 3과 4처럼 선택점에서 서브트리를 교체하는 방식으로 수행된다.

```

begin
t <- 0
initialize P(t) // each individual consists of tree
evaluate P(t)
while (not termination-condition) do
begin
t <- t+1
select P(t) from P(t-1)
recombination P(t) // genetic operation with trees
evaluate P(t)
end
end
    
```

그림 2 GP의 연산 과정  
Fig. 2 Procedure of genetic programming

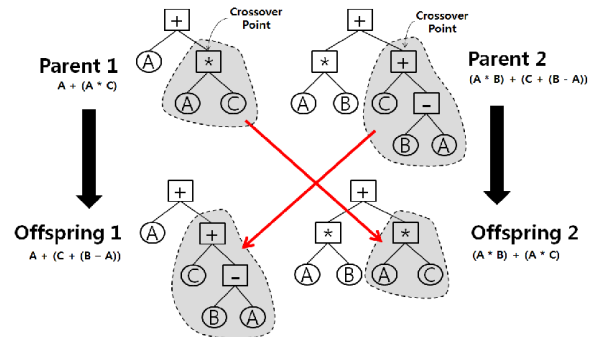


그림 3 GP의 교배 연산  
Fig. 3 Crossover operation of GP

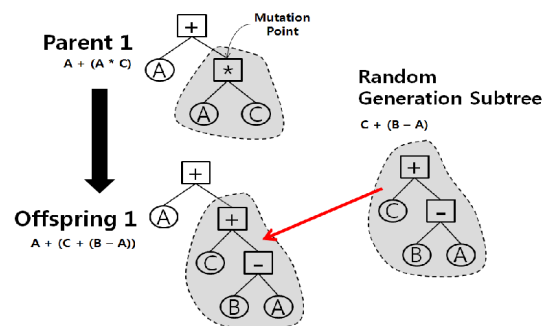


그림 4 GP의 변이 연산  
Fig. 4 Mutation operation of GP

### 3. GP 탐색 문제 개선 - 트리 구성 및 ADF

#### 3.1 트리 개체 구성시 함수 및 터미널 선택

GP는 트리 형태의 유전자가 가변적이며, 구조적으로 열린 공간의 탐색 문제에 적합한 대신에, GA에 비해 방대한 탐색공간을 가진다. 따라서 탐색의 복잡도가 훨씬 높으며, 최적해를 찾기가 더욱 어렵다. 이를 개선하기 위해서 조기 수렴 방지, 전역과 지역탐색의 균형, 개체 트리 크기의 제한, 함수와 터미널 정의 방법, 그리고 트리 구성의 효율성 등의 다양한 접근법이 있으나 [2-11], 본 연구에서는 많이 다루어지지 않는 뒤의 두 가지 측면에 중점을 둔다.

트리 개체의 구성은 함수와 터미널의 조합으로 이루어지는데, 함수는 그림 5, 6에서의 +, - 등의 내부 노드에 위치하며, A, B, C 와 같은 터미널을 자손 노드로 갖는다. 터미널은 말단 노드에 위치하며, 함수의 피연산 인자이다. 함수와 터미널은 문제에 따라 다양하고 제한없이 정의될 수 있다. 널리 쓰이는 함수들의 예가 다음에 나와 있다.

- 사칙연산 : +, -, \*, /
- 초월함수 : sin, cos, exp
- 논리연산 : AND, OR, NOT
- 조건 반복 : if-else, while

터미널은 주로 변수나 상수들이 사용되며, 무인자 함수처럼 터미널 노드에 올 수 있는 함수도 가능하다. 아래 예가 나와 있다.

- 변수 : X, Y
- 상수 : 실수
- 무인자 함수 : move, right, left

해를 구하기 위한 함수와 터미널의 선택에 대한 대표적인 안 내는 Koza[2]의 충분 및 닫힘 (Sufficiency and Closure) 조건으로 설명된다. 터미널과 함수의 집합은 문제를 효율적으로 풀수 있도록 선택되어야 하며, 함수는 다른 함수에 의해서 반환되는 임의의 값을 인자로서 받을 수 있어야 한다. 즉, 범용 반환형을 의미한다.

여기에 형 일치와 (type consistency) 평가 안정성 (evaluation safety) 추가 된다[10]. 형 일치 문제는 서브 트리의 교배나 돌연변이에서 임의의 노드가 선택될 수 있고, 결과로 교배점에서의 모든 교체가 가능하여야 하나, 그래프의 정점과 간선처럼 서로의 형(type)이 호환이 안될 경우, 이에 대한 유전연산을 제한하는 것을 말한다. 평가 안정성은 0으로 나누는 연산과 같이 허용되지 않는 연산으로 인해 프로그램의 수행이 종료되는 것을 피하기 위한 것이다. 0 으로 나눌 경우, 강제로 1을 반환하는 나누기 보호 함수(protected division)를 정의하여 사용한다. 로봇을 이동시키는 move 함수도 벽을 만날 경우는 멈추거나 아무 동작도 하지 않는 기능을 갖는 함수로 수정하여 평가 안정성을 확보한다.

그 외 함수의 선택에 대한 문제에서 수학적 함수의 선택이 진

화과정의 탐색에 영향을 줄 수 있다는 연구가 제시되었다[11]. 여기서 회귀분석의 대상 문제가 sine 함수일 경우, 나누기가 성능에 안 좋은 영향을 끼침을 보였으며, 이를 수정된 나누기 함수로 제거할 수 있음을 보였다. 그러나 대상 문제가 sine 인 경우에만 의존적일 수 있어서 다른 문제에는 미지수이다.

터미널에 대한 선택은 함수의 선택보다는 보다 직관적이지만, 인자들이 많은 경우 인자들이 상관관계에 따른 다중 공선성 문제가 선형회귀분석에서는 흔히 제기되는 문제이다. 그러나, GP에서는 모든 인자가 해마다 다 포함되는 것이 아니므로, 터미널의 집합에는 이용 가능한 대부분의 인자를 포함하고, 진화적 선택에 맡겨두는 방식을 취한다.

문제에 따라 함수와 터미널의 선택이 탐색에 영향을 줄 수 있는 것으로 예상되지만, 아직까지 이들의 구성에 대한 체계적인 연구는 거의 없다. 교배나 돌연변이 수행시에 함수 및 터미널 노드에 대한 선택 확률은 각 0.9와 0.1로 설정하여 함수의 선택 확률을 더 높게 사용하는 것이 일반적이다.

#### 3.2 자동정의함수(Automatically Defined Functions)

ADF[3]는 GP 개체 형태인 트리를 생성시에 유용한 빌딩 블록을 구성하고 이를 다른 곳에서 호출하여 반복적으로 사용하는 서브루틴과 같은 개념이다. 특정 서브트리의 모듈화가 가능하며, 같은 크기의 트리에서 더 복잡화된 연산 수행이 가능하다.

그림 5에 ADF 구성도가 나와 있으며, 왼쪽에는 ADF에 대한 정의부, 오른쪽에는 Result-Producing Branch(RPB)라 불리는 주 함수에 해당하는 부분이 나와 있다[3].

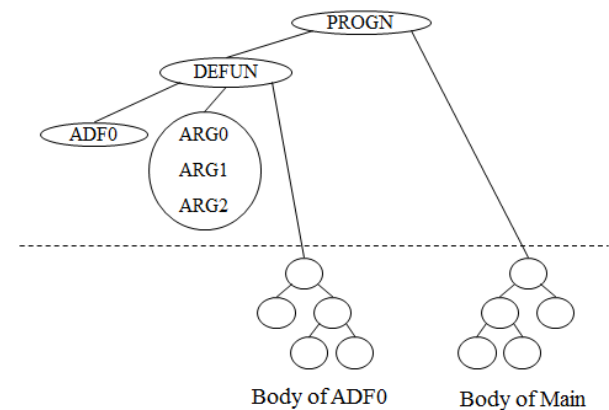


그림 5 ADF 구성도  
Fig. 5 ADF Diagram

ADF의 사용 예가 그림 6에 나와 있다. 그림 6의 윗 부분은 ADF로 정의된 부분이고, 아래 부분은 ADF와 다른 요소들을 조합하여 구성한 주 트리의 예이다. ADF0에 의해서 정의된  $(a+b)/c$  부분식이 아래 주 트리의 ADF0 자리에 치환되어 최종 트리의 식으로 정리된다.

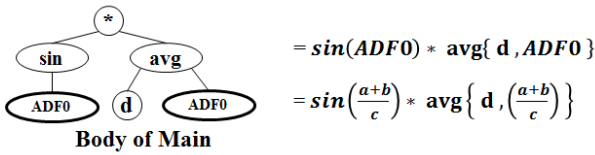
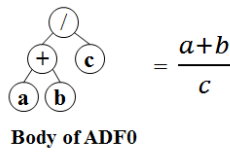


그림 6 ADF의 예  
Fig. 6 Examples of ADF

#### 4. 풍속 예보 보정 기법 응용

##### 4.1 KLAPS 기상 데이터

KLAPS(Korea Local Analysis and Prediction System, 이하 KLAPS)[12]는 한반도 영역에 대해 5km 해상도(235 × 283)의 1 시간 간격으로 수행되는 분석과 예측 시스템으로, 예보 모델 평가시 격자형 구조의 실험자료로 활용된다. KLAPS의 위/경도 자료는 UM의 자료와 동일하게 수직, 수평 5km의 해상도를 가지며, 한반도 9,291개 격자점 중 남한 4,435개 격자점을 대상으로, 육상 4,075개와 산지 360개로 구성되어있다.

본 논문에서는 2007~2009, 2011~2013년 총 6년치의 KLAPS, UM 현업 데이터를 사용하여 예측을 수행한다. 이를 위해, 먼저 시간 단위로 구분되어있는 KLAPS, UM 데이터를 KLAPS 격자점 단위 데이터로 변환하여 실험한다. 그리고 사용하고자 하는 지역통합모델은 하루에 2번(00UTC, 12UTC), 3시간 간격으로 21개의 목표 예보 시간(+06H~+66H)으로 예보 자료를 생성하기 때문에, 관측자료와 예보자료의 동기화 작업이 필요하다. 또한, UM의 64개 예보인자를 기본으로 사용하여 예측식을 생성한다. 그런데, 각 예보인자 별로 나타내고자 하는 지표가 다르고, 상대

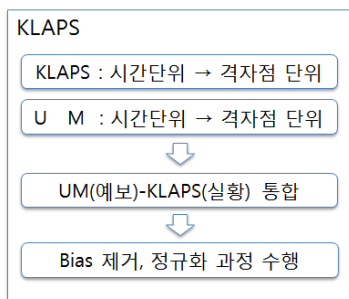


그림 7 풍속 예보를 위한 KLAPS 기상 자료의 재구성 과정  
Fig. 7 Reconstruction procedure of KLAPS weather data for wind speed prediction

적으로 수치적인 크기에서 중요도가 다르게 나타날 수 있기 때문에, 각 예보인자들의 정량적인 크기를 정규화하여 사용한다. 이 과정은 그림 7에 구조도로 설명되어 있다.

##### 4.2 MOS 보정 및 선형회귀분석

MOS(Model Output Statistics)[13]는 수치예보모델의 오차를 줄이는 예보기술로서, 수치예보모델의 예측값에 MOS에 의한 보정량을 합하는 방법이다. 일반적으로, 풍속 예측에 대해서 MOS 기법을 통해 물리적 수치예보모델인 UM[14] 모델의 오차를 보정하는 접근법을 사용한다[15, 16].

기존의 MOS에서 사용되는 예보 보정 기법은 식 (1)과 같이 잠재적 예보인자들과 계수의 곱합을 통한 선형적인 수식으로 구성되어 있다. 여기서 ΔWSS는 풍속에 대한 보정량이며, VAR<sub>i</sub>는 표 1에 나와 있는 UM의 예보인자를 가리킨다. 참고로, UM의 예보인자는 풍속, 풍향, 기온, 습도 등 다양한 기상 인자에 대하여 기압 고도별로 세분화되어 구성된다. 그리고 α는 각 예보인자의 회귀계수이다.

$$\Delta WSS = \alpha_1 VAR_1 + \alpha_2 VAR_2 + \dots + \alpha_N VAR_N \quad (1)$$

그러나 이러한 선형회귀분석은 2.1절의 그림 1에 나타난 바와 같이 기상자료의 비선형적인 특성 분포를 충분히 반영하지 못하고 있다.

##### 4.3 유전프로그래밍 기반 풍속 예보 보정 기법

본 연구자가 제안한 유전프로그래밍(이하 GP) 기반의 풍속 예보를 위한 보정 기법의 개념도가 그림 8에 나와 있다[16].

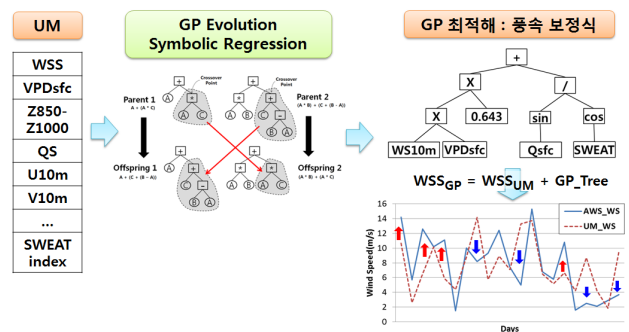


그림 8 GP 기반 풍속 예보보정 기법  
Fig. 8 GP based compensation technique for wind speed forecast

GP 기반 평균 풍속 예보 보정 기법은 UM의 64개 수치 예보 자료를 바탕으로 수행되며, KLAPS 관측 자료의 평균 풍속과의 정량적인 비교 오차(RMSE)를 적합도로 구성한다. 사전 과정으로

GP 트리 생성을 위한 터미널과 함수는 다음과 같이 정의된다. GP 수행에 사용되는 터미널은 표 1에 나온 UM의 64개 예보 인자와 실수 ERC로 이루어지며, 각각의 예보 인자들은 고도에 따라서 1~4개로 구분된다. 각 예보 인자에 대해서는 최소-최대 정규화가 수행된다. GP 함수는 사칙연산(+, -, \*, /), 삼각함수(cos, sine), 평균(avg), 가중필터(wf1, wf2)를 사용한다.

표 1 UM 예보인자  
Table 1 Predictors of UM

Group	Code
Air Temperature	TS, T8, T7, T5
Thickness	DZ18, DZ17, DZ85
Dew-Point Depression	TDD10, TDD8, TDD7, TDD5
Specific Humidity	QS, Q10, Q8, Q7, Q5
Q Difference	DQS5, DQ85, DQ75
Relative Humidity	RHS, RH8, RH7, RH5
Mean RH	MRH17, MRH15, MRH85
Vapour Pressure Deficit	VPDS, VPD10, VPD8, VPD7, VPD5
Zonal Wind	US, U8, U7, U5
Meridional Wind	VS, V8, V7, V6
Wind Speed	WSS, WS8, WS7, WS5
NW Wind Speed	NWS, NW8, NW7, NW5
NE Wind Speed	NES, NE8, NE7, NE5
SW Wind Speed	SWS, SW8, SW7, SW5
SE Wind Speed	SES, SE8, SE7, SE5
Total Rain Amount	PCPN
Lapse rate	LR87, LR85
K-Index	KI
Severe Weather Threatening Index	SWEATI

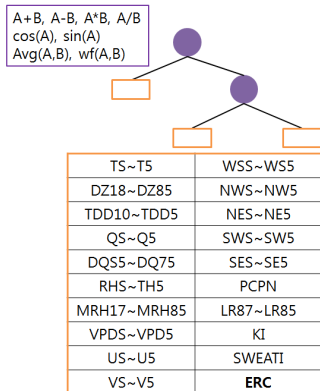


그림 9 GP 개체의 구성  
Fig. 9 Construction of GP individual

2.2절에서 설명한 진화연산의 초기 생성, 평가, 선택 그리고 GP 유전연산의 반복 수행을 거쳐서 각 예보 지점마다 적합도가 가장 우수한 개체를 구한다. 각 개체는 그림 9와 같은 트리의 형태를 가지고 있는데, 원 부분에는 연산함수가, 사각형 부분에는 터미널인 UM 인자가 할당될 수 있음을 나타낸다.

새로운 예보 평균 풍속은 식 (2)와 같이 UM의 예보 평균 풍속과 GP 최적해에 의한 보정식 값을 합한 것이다. 여기서, GP\_Tree는 보정식으로 식 (1)의 ΔWSS에 해당한다.

$$WSS_{GP} = WSS_{UM} + GP\_Tree \quad (2)$$

## 5. 풍속 예보 보정 응용 실험 및 결과

### 5.1 실험 환경 및 설계

KLAPS 격자점에 대한 풍속 예측을 위한 GP 기반의 풍속 보정식 생성은 다음과 같이 수행된다.

- KLAPS 재분석자료 : 2007~2009년, 2011~2013년
- 대상 지점 : 남한 7개 지역 364 개 격자점 (서울 40개, 강원 50개, 충남 40개, 대구 45개, 여수 45개, 부산 43개, 제주도 101개)
- 예보 발표 시각 : 00 UTC
- 목표 예보 시간 : 21개(+06H ~ +66H, 3시간 간격)

각각의 예보 풍속 보정식 생성 실험은 20회 반복 수행하였다. lil-gp[17]를 수정하여 구현하였으며, 실험에 사용한 유전프로그래밍의 파라미터는 다음 표 2와 같다.

표 2 유전프로그래밍 파라미터  
Table 2 Parameters of Genetic Programming

GP Parameters	Values
군집의 크기	200
최대 세대 수	200
최대 트리 깊이	10
초기 트리 구성 방법	half and half
초기 트리 구성 깊이	2 - 3
교배 연산자	확률 0.9, 선택 : 토너먼트(크기 : 7)
돌연변이 연산자	확률 0.1, 선택 : 토너먼트(크기 : 7)

2007~2009, 2011년 4년치 학습 데이터로 구성하고 20번의 GP를 수행한다. 이때 생성된 20개의 해를 2012년 검증(validation) 데이터에 적용한 후, 그 중 최적의 해를 2013년 테스트 데이터에 평가함으로써 과적합 완화를 시도한다(그림 10). 예보 풍속 보정 수식에 대한 성능평가 지표는 RMSE(Root Mean Square Error)



를 사용한다. 비교 대상인 다중선형회귀분석(MLR) 방법에서는 검증 과정이 필요 없으므로 2007~2009, 2011~2012년 5년치를 학습 데이터로 구성하고, 2013년을 테스트 데이터로 선택하여 평가한다. 2010년 데이터는 상당 부분 손실되어 실험에서 제외한다.

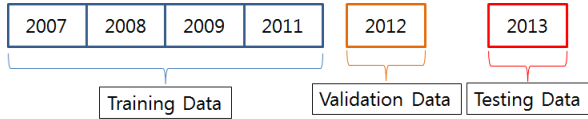


그림 10 GP 학습 데이터 구성

Fig. 10 Composition of learning data for GP

함수 구성의 영향을 분석하기 위해서 일부 지역, 일부 시간대에 대해서 수행한 사전 실험 결과는 표 3과 같으며, 수치는 RMSE를 나타낸다. 사칙연산과 삼각함수(cos, sin) 그리고 평균함수들(avg, wf1, wf2). wf1, wf2 는 서로 가중치가 다른 평균 함수(weighted average)를 의미한다. 실험 결과에서 보듯이 사칙연산만을 사용한 경우보다는 삼각함수와 평균함수들을 함께 사용한 결과가 약간 더 우수하게 나왔으며, 사칙연산과 삼각함수만을 사용한 실험이 가장 우수함을 보였다. 차이는 크지 않지만 어느 정도 영향을 미침을 알 수 있다.

표 3 함수 영향에 대한 사전 실험 결과(UM+GP)

Table 3 Preliminary results of effects for function selection

지역(격자점수) operator	서울 (40)	부산 (43)	제주 (101)	평균 (184)
+, -, *, /	3.5959	1.1473	1.9708	2.1316
+, -, *, /, cos, sin	3.4781	1.1042	1.9068	2.0608
+, -, *, /, cos, sin, avg, wf1, wf2	3.4765	1.104	1.9907	2.1064

본 실험에서는 UM, UM+MLR (UM에 MLR 보정, MLR은 다중선형회귀분석), UM+GP (UM에 GP 보정). 그리고 GP에 ADF 기법을 추가한 보정기법들을 비교한다. GP\_ADF 방법들의 설정은 표 3의 사전 실험에서의 함수 선택 결과를 참고하여 아래와 같이 4가지로 다르게 구성한다.

- GP\_ADF1  
main : +, -, \*, /, sin, cos, avg, wf1, wf2 (64개 인자, ERC)  
ADF0 : +, -, \*, /, (ARG0, ARG1, ARG2)  
ADF1 : sin, cos (ARG3, ARG4)
- GP\_ADF2  
main : +, -, \*, /, sin, cos, avg, wf1, wf2 (64개 인자, ERC)  
ADF0 : +, -, \*, /, sin, cos (ARG0, ARG1, ARG2 )  
ADF1 : +, -, \*, /, avg, wf1, wf2 (ARG3, ARG4, ARG5 )

- GP\_ADF3  
main : +, -, \*, /, sin, cos, avg, wf1, wf2 (64개 인자, ERC)  
ADF0 : +, -, \*, /, sin, cos, avg, wf1, wf2 (ARG 없이 WS (바람) 계열 인자와 ERC만 포함)  
ADF1 : +, -, \*, /, sin, cos, avg, wf1, wf2 (ARG 없이 RH (습도) 계열 인자와 ERC만 포함)
- GP\_ADF4  
main : +, -, \*, /, sin, cos, avg, wf1, wf2 (64개 인자, ERC)  
ADF0 : +, -, \*, /, sin, cos, avg, wf1, wf2 (ARG 없이 WS (바람) 계열 인자만 포함)

각 GP\_ADFi의 설정에서 main 트리는 모두 동일하게 함수에 사칙연산, 삼각함수, 평균함수를 사용하고, 인자로는 64개 UM 인자와 랜덤상수 ERC를 포함한다. 각 ADF의 설정은 함수의 구성 또는 인자의 구성을 다르게 한다. GP\_ADF1, 2, 3은 두 개의 ADF를 사용하며, GP\_ADF4는 하나의 ADF를 사용한다.

GP\_ADF1은 두 개의 ADF에 대해서 각각 사칙연산과 삼각함수로 구성하고, 인자수는 각 3개와 2개로 설정한다. GP\_ADF2는 각 ADF에 사칙연산을 모두 포함시키고, ADF0에는 삼각함수를, ADF1에는 평균함수들을 추가로 포함시킨다. 인자는 둘 다 3개씩 설정한다. GP\_ADF3의 함수는 main과 동일하게 구성하고 인자에 변화를 주어, 바람과 습도 계열 인자들만을 포함시켜, 풍속에 관련된 인자들이 해에 포함되는 비중을 높이도록 한다. GP\_ADF4는 ADF를 한 개로 구성하고 바람 계열 인자만을 포함한다. GP\_ADF3와 GP\_ADF4에서는 인수(ARG)를 사용하지 않는다.

## 5.2 실험 결과 및 분석

도시, 내륙, 해안가, 섬 지역 등 각각의 특성을 고루 반영할 수 있도록, 남한 7개 지역(서울, 강원, 충청, 대구, 전라, 부산, 제주도) 선별하여, 실험을 수행하였으며, 총 364개 격자점에 대한 비교 실험 결과가 표 4에 나와 있다.

UM만의 결과에 비해서 MLR 보정을 한 것이 상당한 성능 향상을 가져왔다. UM 모델이 가지는 오차를 개선한 MOS 보정기법의 근본적인 영향으로 해석된다. GP를 통한 보정은 서울과 강원 지역에서 MLR 보정에 비해서 RMSE 오차가 좀 더 높게 나왔으나, 전체 평균값은 MLR 보정에 비해 약 3.5% 개선되었다.

제안된 4가지 GP\_ADF를 사용한 보정 방법들 모두 MLR 및 GP만의 보정에 비해서 모두 평균값이 우수하게 나왔다. 특히, 두 ADF의 함수 구성을 달리하고 평균값 함수들을 포함한 GP\_ADF2 방법이 전체 평균 성능에서 가장 우수하게 나왔으며, MLR 보정에 비해 약 8.3%, 그리고 GP 보정에 비해서도 약 5.0% 개선되었다. 구체적으로, 서울 지역에서 MLR 보정 및 GP 보정에 비해 대폭 개선되었으며, 강원을 제외한 나머지 6개 지역에서 가장 우수한 결과를 보였다.

참고로, 구성된 보정식의 기상인자 선택에 관한 결과 및 분석은 이전 논문[18]을 참고하기 바란다.

**표 4** 7개 지역에 대한 각 기법의 풍속 예측 성능 비교(RMSE)  
**Table 4** Performance comparison results of wind prediction for 7 areas (RMSE )

00 UTC	서울	부산	제주	대구
UM	4.2583	3.1262	2.3505	1.4656
UM+MLR	3.0350	1.5622	1.9588	1.3065
UM+GP	3.4664	1.0914	1.8917	1.1346
+GP_ADF1	2.8621	1.2794	1.8330	1.1293
+GP_ADF2	2.7959	1.2408	1.8256	1.1273
+GP_ADF3	3.0022	1.2970	1.8402	1.1540
+GP_ADF4	2.9965	1.2974	1.8564	1.1442
00 UTC	충남	강원	여수	평균
UM	2.2146	3.1609	1.7530	3.0549
UM+MLR	2.0016	2.6705	1.4462	2.3301
UM+GP	2.0015	2.6786	1.2312	2.2492
+GP_ADF1	1.9684	2.6211	1.2549	2.1580
+GP_ADF2	1.9655	2.6346	1.2274	2.1362
+GP_ADF3	1.9748	2.6415	1.2791	2.1981
+GP_ADF4	1.9824	2.6552	1.2601	2.1987

여기서 +GP\_ADFi 는 UM + GP\_ADFi 를 의미함

## 6. 결 론

본 논문에서는 자동 모듈화 생성 개념인 ADF를 사용하여 GP 기반 진화적 비선형 회귀분석식 생성 기법을 확장 제안하고, 다량의 인자를 포함하는 풍속 예보의 보정식 생성 문제에 적용하였다. 사례 연구로서 서울, 부산, 제주도를 포함한 남한의 7개 지역, 364개 격자점에 대해서 2007년~2009, 2011~2013년의 방대한 기상 데이터를 대상으로 다양한 최적화 실험을 수행하고, 풍속 예측에 대한 성능을 UM 및 다중선형회귀분석법과 비교하였다.

제안된 ADF 기반 보정 방법들이 모두 MLR 및 GP만의 보정에 비해서 우수한 성능을 보였으며, 가장 우수한 GP\_ADF2 방법이 전체 평균 성능에서 MLR 보정에 비해 약 8.3%, 그리고 GP 보정에 비해서도 약 5.0% 개선되었다.

향후, ADF 구성 방법의 체계화를 통해 비선형 회귀식 생성 기법의 성능을 높이고, 실험을 남한 전체로 확장하여 수행할 계획이다.

## References

[1] D. E. Goldberg, Genetic Algorithms in Search, Optimi-

zation, and Machine Learning, Addison-Wesley, Reading, MA, 1989.

[2] J. R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection, The MIT Press, 1992.

[3] J. R. Koza, 1994 : Genetic Programming II: Automatic Discovery of Reusable Programs, The MIT Press, 1994.

[4] J. R. Koza, F. H. Bennett, D. Andre and M. A. Keane, III, Darwinian Invention and Problem Solving, Morgan Kaufmann Publishers, USA, 1999.

[5] K. Seo S. Hyun E. D. Goodman, "Genetic Programming-Based Automatic Gait Generation in Joint Space for a Quadruped Robot," Advanced Robotics, Vol. 24, No. 15, pp. 2199-2214, 2010.

[6] S. Luke, "Issues in Scaling Genetic Programming: Breeding Strategies, Tree Generation, and Code Bloat," PhD of University of Maryland, 2000.

[7] R. Poli and J. Page, "Solving High Order Boolean Parity Problems with Smooth Uniform Crossover, Sub Machine Code GP and Demes," Genetic Programming and Evolvable Machines, Volume 1, Issue 1/2, pp.37-56, April 2000.

[8] Nguyen, X. Hoai, B. McKay and D. Essam, "Representation and structural Difficulty in Genetic Programming," Evolutionary computation, IEEE Transactions on Volume 10, Issue 2, pp.157-166, April 2006.

[9] K. Seo, C. Pang, "Tree-Structure-Aware Genetic Operators in Genetic Programming", JEET(Journal of Electrical Engineering and Technology), vol.9, no.2, pp.755-761, March 2014.

[10] R. Poli, W. B. Langdon, N. F. McPhee, A Field Guide to Genetic Programming, Lulu Enterprises., 2008.

[11] W. Gang and T. Soule, "How to Choose Appropriate Function Sets for GP," Genetic Programming 7th European Conference, EuroGP 2004, Proceedings, LNCS, Vol. 3003, pp. 198-207, Springer-Verlag, 2004.

[12] Korea Meteorological Administration, <http://www.kma.go.kr>.

[13] H. R. Glahn, D. A. Lowry, "The use of model output statistics (MOS) in objective weather forecasting", J. Appl. Meteor., 11, pp. 1203 - 1211, 1972.

[14] United Kingdom Met Office's website : <http://www.metoffice.gov.uk>

[15] K. Seo, B. Hyeon, S. Hyun, and Y. Lee, "Genetic Programming-Based Model Output Statistics for Short-Range Temperature Prediction", Lecture Notes in Computer Science, Springer-Verlag, Vol. 7835, pp. 122 - 131, 2013

[16] B. Hyeon, K. Seo, Y. Lee, "Evolutionary Nonlinear Regression Based Compensation Technique for Short-range

Prediction of Wind Speed using Automatic Weather Station”,  
The Transactions of the Korean Institute of Electrical  
Engineers, vol. 64, No.1, pp. 107-112, 2015

- [17] D. Zongker and B. Punch : Lil-GP User’s Manual, Michigan  
State University, 1995.
- [18] D. Kim, and K. Seo, “Comparison of Linear and Nonlinear  
Regressions and Elements Analysis for Wind Speed  
Prediction”, Journal of Korean Institute of Intelligent  
Systems, Vol. 25, No. 5, pp. 477-482, 2015

---

## 저 자 소 개



### 오 승 철 (Seungchul Oh)

2014년 서경대학교 전자공학과 졸업(학 사).  
2014년~현재 서경대학교 대학원 전자공학  
과 석사과정. 관심분야는 진화연산, 기상예  
측, 영상인식.



### 서 기 성 (Kisung Seo)

1993년 연세대학교 대학원 전기공학과 졸업  
(박사). 1999~2003년 Michigan State Uni-  
versity, Genetic Algorithms Research and  
Applications Group, Research Associate,  
2002~2003년 Michigan State University,  
Electrical & Computer Engineering, Visi-  
ting Assistant Professor, 2011~2012년 Michigan State  
University, BEACON(Bio/computational Evolution in Action  
CONsortium) Center, Visiting Scholar, 1993년~현재 서경대  
학교 전자공학과 부교수. 관심분야는 진화연산, 기상예측, 지능  
로봇.