

풍속 예측 보정을 위한 Genetic Programing 탐색 기법의 개선

오승철*, 서기성*
서경대학교 전자공학과*

Improvement of Search Method of Genetic Programing for Wind Prediction MOS

Seungchul Oh*, Kisung Seo*
Dept. of Electronics engineering, SeoKyeong University*

Abstract - 풍속은 다른 기상요소들보다 순간 변동이 심하고 국지성이 강하여 수치 예보 모델만으로 예측의 정확성을 높이기 어렵다. 기상청의 단기 풍속 예보는 전 지구적 통합 예보모델인 UM(Unified Model)의 예측값에 MOS(Model Output Statistics)를 통한 보정을 수행하며, 보정식의 생성에 다중선형회귀분석 방법을 사용한다. 본 연구자는 유전프로그래밍(Genetic Programming)을 이용한 비선형 회귀분석 기반의 보정식 생성을 통하여 이를 개선한 바 있는데, 본 연구에서는 보다 향상된 성능을 얻기 위하여 GP 기법 측면에서 Automatically Defined Functions과 다군집(Multiple Populations) 수행을 통해 성능을 높이고자 한다.

1. 서 론

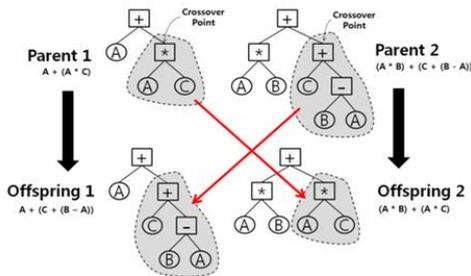
기상은 불확실성과 혼돈성이 큰 자연계의 현상으로서, 이 중에서 풍속은 다른 기상요소들보다 순간변동이 심하고, 국지성이 강하여 예측이 더욱 어렵다. 현재 기상청의 단기 풍속 예보는 영국에서 개발된 통합모델인 UM(Unified Model) 예보 모델의 예측값에 선형회귀분석 기반의 MOS(Model Output Statistics)에 의한 보정값을 합하는 방법을 사용하고 있다[1]. MOS는 통계적 분석을 통해 수치예보 모델이 가지는 체계적인 오차를 보정하는 방법이다. 그러나 선형적 보정만으로는 기상요소의 복잡한 비선형 특성을 반영하기 힘들다.

이를 개선하기 위하여 이전 연구에서[2,3] 진화적 기호회귀 분석기법인 유전 프로그래밍(Genetic Programing, 이하 GP)을 사용하여 풍속 예보에 대한 비선형 보정 수식을 생성하는 기법을 제안하였다. 본 연구에서는 GP 기법에 ADF(Automatically Defined Functions)와 다군집 탐색을 추가하여 성능을 비교한다. 서울, 부산, 제주 세 지역에 대해서, 2007년~2013년의 KLAPS 데이터를 대상으로 보정식 최적화 실험을 수행하였다.

2. GP 기반의 풍속 예보 보정식 생성

2.1 Genetic Programing

유전 프로그래밍 GP는 유전자가 트리로 구성된 진화연산의 일종으로 구조적으로 제한되지 않은 해의 생성이 가능하고, 기호적(symbolic) 회귀 분석 등이 주요 응용 사례이다[4]. 유전 프로그래밍은 트리의 터미널과 함수가 진화 연산에 의하여 자연 선택되며, 고차원적인 수식의 표현이 가능하므로, 비선형적인 예보 보정수식을 생성하는데 적합하다. 예보 풍속을 보정하기 위하여 UM의 64개 예보인자와 사칙연산, 삼각함수와 같은 연산자를 트리 형태로 구성한 비선형 수식을 생성한다(그림 1).

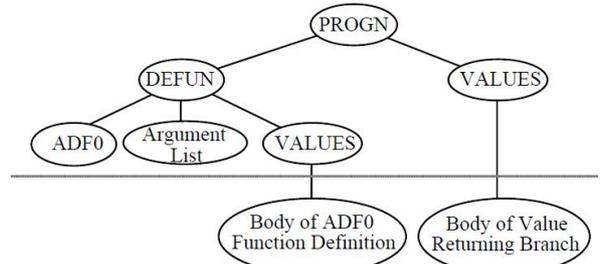


<그림 1> GP 연산 구조도

2.2 Automatically Defined Functions

ADF[5]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 복잡한 연산 수행이 가능하다.

그림 2에 ADF 구성도가 나와 있으며, 왼쪽에는 ADF에 대한 정의부, 오른쪽에는 Result-Producing Branch(RPB)라 불리는 주 함수에 해당하는 부분이 나와있다[5].

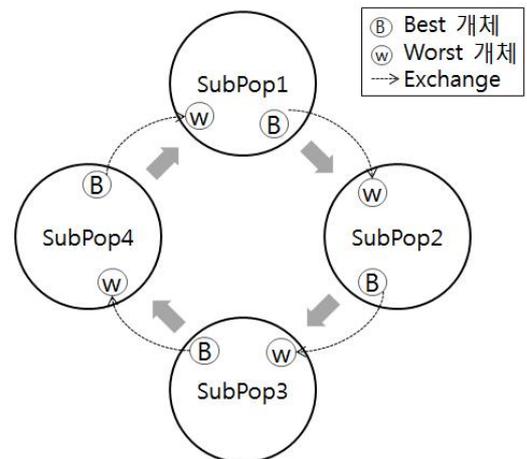


<그림 2> ADF 구성도

2.3 다군집 탐색

다군집(Multiple Populations) 탐색은 기존의 단일 군집을 몇 개의 부분 집단(subpopulations)으로 나누어 탐색을 수행하는 기법으로서, 격리를 통해 해의 다양성 유지하여 조기 수렴을 방지하는 목적을 가진다. 나누어진 부분 군집의 개체수가 적기 때문에 발생하는 탐색 성능의 저하를 일정 세대마다 군집간의 이주를 통해 보완한다.

이 때, 한 군집의 일부 열성(worst) 개체들을 다른 군집의 우성(best) 개체들로 교체한다. 이를 통해, 기존의 개체들이 다른 군집에서 유입된 새로운 개체들과 경쟁을 통해 진화해나가는 방법이다. 다군집 구성은 그림 3과 같이 고리(Ring)형을 사용한다.



<그림 3> 고리구조의 다군집 탐색

3. 실험 및 결과

본문 2.1에서 언급한 ADF와 multiple populations를 적용하여 서울, 부산, 제주 세 지역에 대해서, 2007년~2013년의 KLAPS 데이터를 대상으로 최적화 실험을 수행하였다.

3.1 실험환경

제안된 기법에 의한 풍속 예측 실험은 KLAPS(실황자료) 데이터에 대해서, 서울, 부산 그리고 제주도를 대상으로 수행되었으며, 각 지역은 수도권 내륙인 서울, 바다와 인접한 지역인 부산, 그리고 섬 지역인 제주도로 지역적 특성이 서로 다른 지역들이다. 실험환경은 다음과 같다.

- KLAPS 재분석자료 : 2007년~2013년
- 격자점 수(KLAPS) : 184 개 격자점
(서울 40개, 부산 43개, 제주도 101개)
- 예보 구간 : 21개(+06H ~ +66H)
- 발표 시각 : 00UTC

실험 방법은 각 지역 격자점에 대해 ADF만 적용한 실험 1, 실험 2와 ADF 및 다군집을 동시에 적용한 실험3, 실험4으로 구분하고, UM에 GP 보정을 한 이전의 실험 결과와 비교하였다.

실험 1		
RPB	function	사칙연산(+, -, *, /), 삼각함수(sin, cos), 평균함수 avg, 필터(wf1, wf2)
	terminal	UM 64개 인자, 랜덤상수
ADF0	function	사칙연산(+, -, *, /), 삼각함수(sin, cos), 평균함수 avg, 필터(wf1, wf2)
	terminal	UM 64개 인자 중 바람 관련 인자 12개
실험 2		
RPB	function	사칙연산(+, -, *, /), 삼각함수(sin, cos), 평균함수 avg, 필터(wf1, wf2)
	terminal	UM 64개 인자, 랜덤상수
ADF0	function	사칙연산(+, -, *, /)
	terminal	ARG0, ARG1, ARG2
ADF1	function	삼각함수(sin, cos)
	terminal	ARG3, ARG4
실험 3		
실험 1과 동일한 ADF 설정 + Multiple populations		
실험 4		
실험 2과 동일한 ADF 설정 + Multiple populations		

다군집 수행을 위한 파라미터 설정은 다음과 같다.

- 형태 : Ring 형
- subpopulations : 4개 군집
- 교체 빈도 : 50세대
- 교체 개체 수 : 5개체

3.2 실험결과

Multiple populations를 적용하지 않고 ADF에 대한 설정만 적용해 준 실험을 실험환경에서 설정한 대로 실험1, 실험2를 수행하였다. 그리고 ADF와 Multiple populations를 동시에 적용한 실험 중 실험1의 ADF 설정과 Multiple populations를 설정한 실험을 실험3, 실험2의 ADF 설정과 Multiple populations를 설정한 실험을 실험4로 수행하였다. 네 가지 실험과 기존의 비선형회귀분석을 통한 GP 보정 풍속 예측 기법과[3] 비교한 결과가 표 1에 나와 있다.

〈표 1〉 ADF만 적용, ADF와 multiple populations의 동시 적용과 기존 GP 보정 풍속 예측과의 성능 비교 결과

	서울	부산	제주	평균
UM+△GP	3.4664	1.0914	1.8917	2.0470
실험1	2.9965	1.2974	1.8564	1.9736
실험2	2.8621	1.2794	1.8330	1.9273
실험3	2.9366	1.2212	1.8540	1.9414
실험4	2.8010	1.2221	1.6002	1.8940

표 1에서 보듯이, ADF 1개를 적용한 실험 1은 이전의 GP 보정 실험에 비해 부산지역은 RMSE 성능이 약 0.13정도 나빠졌으나, 서울 지역에서 약 0.46 정도 개선되었으며, 제주 지역에서도 미미하게 개선된 점을 확

인 할 수 있다. ADF 2개를 적용한 실험 2는 기존의 GP 보정 실험과의 비교에서 실험1과 비슷한 추세이고, 실험 1에 비해서 서울, 부산, 제주 세 지역에서 모두 약간씩 개선된 것을 확인할 수 있다.

실험 1의 ADF에 다군집 탐색을 적용한 실험 3에서도 실험 1, 실험 2와 유사한 결과를 보였다. 실험 2의 ADF에 다군집 탐색을 적용한 실험 4에서는 기존의 GP 보정 실험에 비해 서울 지역의 RMSE가 약 0.67 정도 개선되었고, 부산 지역은 0.13 정도 저하되었지만, 제주 지역은 약 0.3정도 개선되어 실험 방법중에서 우수함을 확인할 수 있었다.

4. 결 론

풍속 예측을 위한 보정식 생성 문제에 대해서 개선된 GP 탐색 기법과 이전의 GP 비선형 보정기법을 비교하였다. KLAPS 데이터에 대한 서울, 부산, 제주지역 184개 격자점을 대상으로 세 지역 평균 풍속 예측 보정기법의 비교 실험 결과, ADF와 다군집 탐색을 적용한 실험 4의 RMSE가 1.894으로 이전의 GP 보정 기법에 의한 RMSE 2.047에 비해 0.153, 약 7.5%만큼 개선되었다.

특히, 지역별로 RMSE 오차가 크게 나타났던 서울 지역에서 0.67만큼 크게 개선된 점을 미루어볼 때 RMSE 오차가 큰 다른 지역에서도 유사한 효과를 볼 것으로 예상된다. 향후 남한지역 전체의 격자점 4435개에 대해서 제안된 기법을 적용하여 이전의 결과를 향상시키고자 한다.

[참 고 문 헌]

- [1] 추성호, 이용희, 안광득, 정관영, “하모니 탐색 알고리즘을 이용한 한반도 바람 예측 모델 개발”, 한국지능시스템학회 2013 춘계학술대회 논문집, 제 23권, 제 1호, pp. 198-199, 2013
- [2] 현병용, 서기성, 이용희, “강풍 예측을 위한 진화적 회귀분석 기반의 단기 풍속 예보 보정 기법”, 정보 및 제어 학술대회 CICS'2013 논문집, pp. 76-77, 2013. 10. 17-19
- [3] 오승철, 서기성, “진화적 풍속 예측 모델 생성 기법 및 KLAPS 재분석 자료 실험”, 정보 및 제어 심포지엄, 2015
- [4] Koza, J. R., 1992: Genetic Programming: On the Programming of Computers by Means of Natural Selection, The MIT Press, 1992.
- [5] Koza, J. R., 1994: Genetic Programming II: Automatic Discovery of Reusable Programs, The MIT Press, 1994.