

진화적 비선형 보정 및 SVM 분류에 의한 강풍 특보 예측 기법

Evolutionary Nonlinear Compensation and Support Vector Machine Based Prediction of Windstorm Advisory

서 기 성*
(Kisung Seo)

Abstract - This paper introduces the prediction methods of windstorm advisory using GP nonlinear compensation and SVM. The existing special report prediction is not specialized for strong wind, such as windstorm, because it is based on the wide range of predicted values for wind speed from low to high. In order to improve the performance of strong wind reporting prediction, a method that can efficiently classify boundaries of strong wind is necessary. First, evolutionary nonlinear regression based compensation technique is applied to obtain more accurate values of prediction for wind speed using UM data. Based on the prediction wind speed, the windstorm advisory is determined. Second, SVM method is applied to classify directly using the data of UM predictors and windstorm advisory. Above two methods are compared to evaluate of the performances for the windstorm data in Jeju Island in South Korea. The data of 2007-2009, 2011 year is used for training, and 2012 year is used for test.

Key Words : Windstorm, prediction, GP nonlinear compensation, SVM, AWS, UM, KLAPS

1. 서 론

바람은 대기의 운동에 의한 물리적 현상으로 기압, 열순환, 위도, 지형, 계절의 영향을 받는다. 그리고 규모에 따라 국지적 바람, 중규모 대류계, 대규모 대류계 등으로 분류된다. 기상 예측에서 중요시하는 강풍은 자연 재해를 발생시킬 수 있는 수준을 의미한다[1].

강풍의 종류로는 태풍, 저기압 전선풍, 계절풍 등이 있는데, 태풍은 해양에서 발생하여 내륙까지 영향을 미치는 대표적인 사례이다. 강풍의 빈도수는 호우보다 많으나 지속시간이 짧은 순간적인 특성이 있다[1]. 특히, 일정 속도 이상의 바람이 불 때 발효되는 강풍 특보의 경우는 수가 적고, 기존의 축적된 자료가 충분치 않아서 예측이 어렵다[2].

기상 예측 연구 중 바람에 관련된 것은 풍속 예측이 주를 이루고 있으며, 관련 연구 동향은 다음과 같다. Sweeney 등은 다양한 통계적인 후처리 방법을 비교하고, 이들을 MSE 기반의 예측기법으로 결합하여 풍속 예측의 오류를 감소시키는 연구를 수행하였다[3]. Zhang 등은 WTT 기법과 SAM 접근방법, RBFNN의 하이브리드 모델 기반의 단기 풍속 예보기법을 제안하였다[4]. 또한, Wang 등은 딥러닝 기법중의 하나인 deep belief network(DBN)과 웨이블릿 변환을 사용하여 기본 풍속 데이터를 다른 주파수로 분해하여 특성을 개선시킨 후 이를 풍속 예측에

사용하였다[5].

그러나 이들 풍속 연구들은 시계열 데이터에 대한 예측으로서 다수의 기상 인자를 고려하지 않고 있다.

한편 대규모의 기상 인자를 사용하여 풍속을 예측한 연구들은 다음과 같다. Seo 등은 진화적 기호회귀 분석기법인 Genetic Programming (GP)[6]을 사용하여 비선형 보정 수식을 생성하는 기법을 제안하고, 풍속 예측에 대한 실험을 수행하였으며, UM 모델의 결과와 비교했다[7]. 또한 풍속 예측용 보정 수식 생성에 관한 또 다른 연구에서 다양한 다중선형회귀분석법과 유전 프로그래밍 기반의 비선형 접근법의 성능을 비교한 바 있다[8].

이들 연구는 풍속 예측을 위주로 한 접근들로서, 풍속 데이터의 대부분을 차지하는 중속 이하의 풍속을 예측 하는데 적합하다. 일반적인 풍속의 예측값에 기초한 기존의 특보 예측은 풍속의 범위가 넓어서 특보에 특화되어 있지 않다. 따라서 강풍 특보 예측의 성능을 높이기 위해서는 강풍의 경계 부분을 효율적으로 분류할 수 있는 기법이 필요하다.

본 연구에서는 GP 기반의 비선형 보정 방식의 풍속 예측 접근법[7]과 SVM (Support Vector Machine)[9] 기반의 기계학습 방법을 사용한 특보 예측 기법을 비교한다. 제주지역의 101개 격자점에 대해서, 2007년~2009, 2011년의 UM[10] 및 KLAPS [2] 데이터를 대상으로 학습을 수행하고, 2012년의 데이터를 대상으로 성능 평가를 수행한다.

2. 강풍 특성 및 AWS, UM 데이터

강풍 특보는 주의보와 경보로 나뉘며, 풍속 14 m/s, 순간풍속

* Corresponding Author : Department of Electronics Engineering, Seokyeong University, Korea
E-mail: ksseo@skuniv.ac.kr

Received : November 13, 2017; Accepted : November 16, 2017

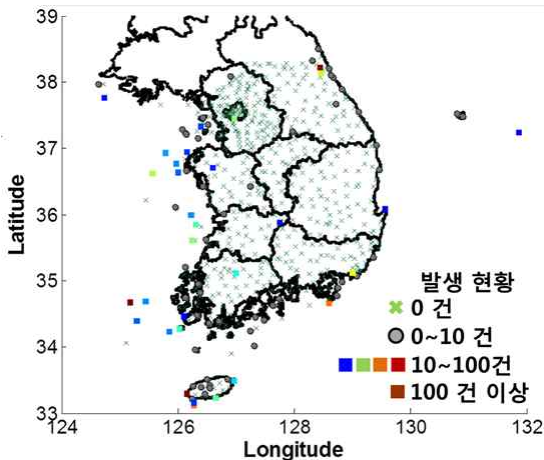


그림 1 강풍특보 발생 건수

Fig. 1 The number of windstorm occurrences

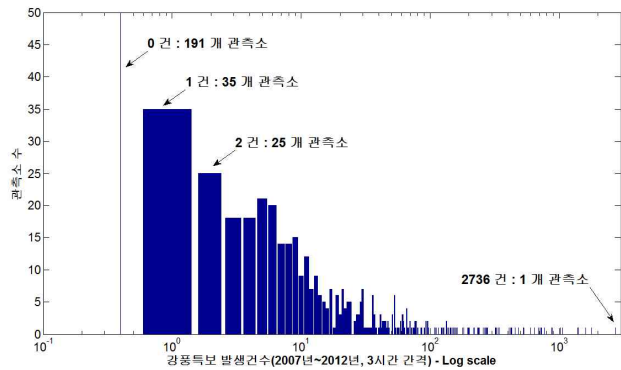


그림 2 강풍특보 발생 건별 관측소 수

Fig. 2 The number of AWS for each occurrence numbers of windstorm

20 m/s 이상일 때 강풍 주의보가, 풍속 21 m/s, 순간풍속 26 m/s 이상이 예상될 때 강풍 경보가 발효된다. 호우는 6시간 누적량을 기준하나, 강풍은 순간적 시점을 기준으로 한다. 참고로, 바람의 90% 이상은 9m/s 미만의 중저속이다. 그림 1은 2007~2012년(2010년 제외) 사이에 남한 지역의 자동기상관측소(Automatic Weather Station, 이하 관측소 또는 AWS)[2]에서 강풍특보가 발생한 건수를 나타내며, 평균풍속을 기준으로 3시간 간격 마다 측정한 것이다. AWS는 기온, 습도, 평균 풍속, 순간 풍속 등의 기상 인자들을 센서와 컴퓨터를 통해 자동으로 관측하고 저장하는 곳이다.

그림 1에서 대부분의 강풍은 남해안 지역과 제주도에 몰려있다. 내륙지방에서는 5년동안 한 건도 발생하지 않은 지역이 대다수임을 알 수 있다.

그림 2는 같은 기간에 대해서 강풍특보 발생 건별 관측소의 수를 나타낸 것이다. 총 615개 관측소중 절반이 5건 이내의 강풍 발생 빈도를 나타낸다. 또한 전체 풍속 데이터에 대해서 강풍 데이터가 차지하는 비율은 0.8% 미만으로 현저히 적다[2]. 따라서

모든 데이터를 사용하여 강풍 특보를 예측할 경우, 예측 모델이 데이터의 대부분을 차지하는 중저속 데이터에 의존하여 학습되기 때문에 강풍특보에 특화된 예측 모델을 생성하기가 어렵다. 그러므로 이러한 데이터의 특성을 고려하여, 강풍 데이터와 비강풍 데이터의 비율을 적절히 설정해야 한다.

수치예보모델인 UM은 풍속, 풍향, 기온, 습도 등 다양한 기상 인자에 대하여 목표 고도별로 세분화된 수치 예보 자료를 생성한다. 본 논문에서는 UM의 64개 예보인자를 활용하여 예보 풍속의 정확도를 향상시키는 작업을 수행한다. 그런데, 각 예보인자 별로 나타내고자 하는 지표가 다르고, 상대적으로 수치적인 크기에서 중요도가 다르게 나타날 수 있기 때문에, 각 예보인자들의 정량적인 크기를 정규화하여 사용한다.

표 1 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

표 1은 현재 기상청에서 사용되는 수치 예보 모델인 UM(Unified Model)의 64개 예보인자와 해당 코드를 나타낸 것이다. 각 그룹의 예보인자명 뒤에 붙는 기호들의 의미는 다음과 같다. 처음 나온 S는 표준(standard), 8, 7, 5는 각 850hPa, 700hPa, 그리고 500hPa를 가리킨다. 그리고 두자리 숫자로 된 85는 850-500hPa, 87은 850-700hPa, 75는 700-500hPa, 17은 1000-700hPa, 15는 1000-500hPa 사이의 평균값을 의미한다. 예로, 표 1의 첫줄에 나온 Air Temperature 그룹에서 TS는 지표면 2m에서의 온도, T8은 850hPa, T7은 700hPa, 그리고 T5는 500hPa에서의 온도를 의미한다. 여기서, 850hPa는 대략 1,500m 높이에 해당한다.

사용하고자 하는 UM 지역모델은 하루에 2번(00UTC, 12UTC), 3시간 간격으로 21개의 목표 예보 시간(+06H~+66H)으로 예보 자료를 생성하기 때문에, 관측자료와 예보자료의 시간적인 동기

화 작업이 필요하다.

또한, 설치되어 있는 지점 단위로 구분되어있는 AWS 데이터와 격자점으로 구성되어있는 KLAPS(재처리자료) 및 UM 데이터 간의 대응 변환이 필요하다.

3. GP와 SVM 기반의 강풍 특보 예측 기법

3.1 GP 보정 기반 특보 예측

UM 모델은 전지구 모델과 지역 모델을 포함한다. 현재 사용하는 지역모델은 한반도 지역의 기상 특성을 충분히 반영하지 못하고 있기 때문에 이를 보정한 방식이 사용되고 있다. 본 연구에서는 보정 방식중 선형회귀분석의 단점을 보완한 진화적 비선형 보정법을 먼저 적용한다. 이를 위해서 지난 연구에서[7] 제안된 GP 기반 진화적 비선형 보정기법을 사용하여 풍속 예측 값을 예측하고, 이 결과를 가지고 특보 여부를 분류한다. 그림 3에 이 과정이 나와 있다.

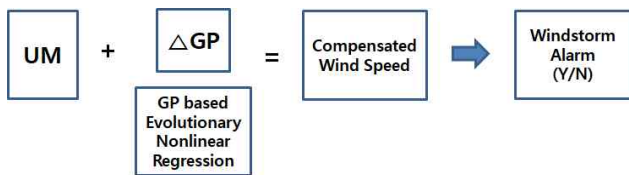


그림 3 GP 비선형회귀에 의한 UM 풍속 보정 및 특보 분류
 Fig. 3 GP based nonlinear compensation of UM wind speed and its classification of windstorm advisory

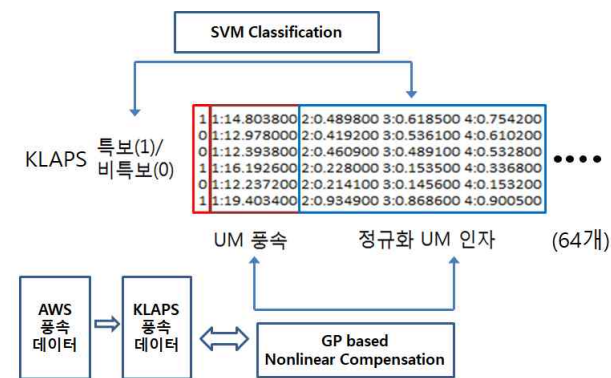


그림 4 강풍 특보 데이터 구성
 Fig. 4 Construction of windstorm data

그림 4에 나와 있는 강풍 특보 데이터를 예로 들어 구체적으로 설명하면 다음과 같다. 정규화된 UM 인자를 가지고 수치 모델인 UM 풍속값을 실제 풍속값인 KLAPS 풍속 값에 근접하게 예측할 수 있도록 학습 데이터를 사용하여 기호회귀분석 기반의 GP 비선형 보정식을 최적화한다(그림 4 하단).

3.2 SVM 기반 특보 예측

SVM[8]은 기계학습의 한 종류로서, 선형 및 비선형 분류가 가능하고 최외각 벡터(support vector)간의 여유를 최대화하는 알고리즘이다. 즉, 선형 및 비선형 회귀분석법등에 의한 풍속 예측값으로 단순 판별하던 방식과는 달리, UM 인자와 KLAPS의 강풍 특보 발생 여부를 SVM 기반의 학습으로 예측한다. 그림 5에 과정의 설명이 나와 있다.



그림 5 SVM based 강풍 특보 분류
 Fig. 5 SVM based classification of windstorm advisory

구체적으로 그림 4의 상단에서와 같이 KLAPS 특보 데이터와 UM의 정규화 인자 데이터 집합만을 가지고 특보와 비특보 데이터를 분리하여 학습 및 테스트 데이터 집합을 구성한다.

3.3 강풍 특보 분류를 위한 학습 데이터 구성

강풍 특보 분류에 효율적인 학습 집합 구성을 위해서, 특보 풍속과 비특보 풍속에 대한 언더 샘플링 비율을 다양하게 변화시켜 분석한다. 학습을 위한 풍속 데이터는 그림 6과 같이 구성한다.

그림 6-(1)은 학습시킬 풍속 데이터를 KLAPS 풍속을 기준으로 고속에서부터 저속까지 정렬시킨 것을 나타낸다. 그림 6-(2)는 정렬된 데이터를 강풍, 고속, 중속, 저속 4단계로 나누고, 강풍(S) : 고속(H) : 중속(M) : 저속(L)의 비율로 학습 집합을 구성하는 것을 나타낸다.

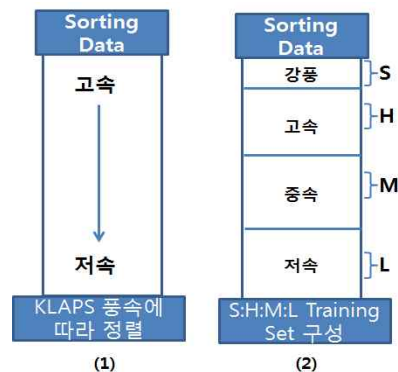


그림 6 강풍 데이터의 샘플링 방법
 Fig. 6 Sampling method for windstorm data

강풍은 KLAPS가 14m/s 이상인 데이터이며, 고속풍은 강풍을 제외한 데이터의 상위 1/3의 풍속 데이터이고 중속풍은 2/3, 저속풍은 그 이하의 데이터이다. S:H:M:L의 비율은 학습시킬 데이터의 강풍의 개수에 따라 비례한다. 예를 들어 S:H:M:L이 1:1:1:1인 경우는 모든 강풍 데이터를 학습 집합에 포함시키고

고속, 중속, 저속 데이터는 강풍 데이터의 개수만큼 학습 집합에 포함 시킨다는 것을 의미한다. 강풍의 수에 비해 고속, 중속, 저속 데이터가 매우 많기 때문에, 랜덤한 인덱스의 풍속 및 기상인자가 선택되도록 샘플링을 수행한다.

4. 실험 및 결과

4.1 실험 방법 및 환경

GP 기반의 비선형 보정 방식의 특보 예측 기법과 SVM 기반의 특보 예측 기법을 다음 환경에 대해서 비교한다. 실험 대상 지역과 예보 시간, 사용된 데이터에 대한 정보는 다음과 같다.

- 대상 지점 : 제주도 지역 101개 격자점
- 예보 발표 시각 : 2개(12 UTC)
- 목표 예보 시간 : 21개(+06H ~ +66H, 3시간 간격)
- 학습 데이터 : 2007~2009년,
2011년 테스트 데이터 : 2012년

강풍 특보 예측의 성능평가를 위해서 표 2의 혼동행렬(confusion matrix)을 사용하였으며, 식 (1)-(4)에 정의된 CSI, PAG, ACC, ETS 지표를 사용하였다.

표 2 관측 및 예보 혼동 행렬

Table 2 Confusion matrix representing event observation and forecast

		Event Observation		
		Yes	No	소계
Event Forecast	Yes	TP	FP	TP+FP (Forecast Yes)
	No	FN	TN	FN+TN (Forecast No)
	소계	TP+FN (Observed Yes)	FP+TN (Observed No)	TP+FN+FP+TN

$$CSI = \frac{TP}{TP + FN + FP} \tag{1}$$

$$PAG = \frac{TP}{TP + FP} \tag{2}$$

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \tag{3}$$

$$ETS = \frac{TP - a_r}{TP + FN + FP - a_r}, \quad a_r = \frac{(TP + FN)(TP + FP)}{TP + FN + FP + TN} \tag{4}$$

4.2 실험 결과

SVM 기반 강풍 특보 예측 기법을 중심으로 UM 모델과 GP 보정 기법과 비교 하였다. SVM1~SVM5는 학습 데이터 집합의 풍속 구성 비율을 변화 시킨 것이며 구성비는 표 3과 같다.

표 3 SVM 학습 집합의 구성 비

Table 3 Ratio of SVM Training Set

	S	H	M	L
SVM1	1	1.3	0.5	0
SVM2	1	1	2	2
SVM3	1	1	0.3	0.7
SVM4	1	1	0.5	1
SVM5	1	1	1	1
SVM4	1	1	2	0.5
SVM5	1	1	0.7	1.3

학습 집합 구성비를 변화시키면서 실험을 진행하였고 구성비가 변화함에 따라 예측 결과의 양상이 달라지는 것을 확인하였다. 표 4는 강풍 특보 예측에 대한 사건 빈도수를 나타낸다. 굵은 숫자는 성능이 가장 우수한 경우를 표시하였다. SVM 2에서 False Alarm과 Negative Correct가 가장 우수하게 나왔고, SVM 3에서 Hits와 Miss가 가장 좋은 결과를 보였다.

표 4 강풍 특보 예측에 대한 사건 빈도수

Table 4 Results of event numbers for windstorm prediction

	UM	GP	SVM1	SVM2	SVM3	SVM4	SVM5
Hits	2391	708	699	316	2844	2636	879
Miss	3268	4951	4960	5343	2815	3025	4780
False Alarm	12591	1254	3368	321	13244	11534	2922
Negative Correct	749552	760889	758775	761822	748899	750609	759221

표 5에는 4가지 성능 평가지수에 대한 결과가 나와있다. 역시 굵은 숫자가 가장 우수한 성능을 나타낸다. SVM 2에서 PAG와 ACC가 가장 우수하게 나왔고, SVM 4에서 CSI와 ETS가 가장 좋은 결과를 보였다.

구체적으로 표 5에서 SVM2와 같이 강풍의 비율에 비해 고, 중, 저속풍의 비율이 높을 경우, Hit는 적지만 False Alarm의 수가 현저히 줄어들어 PAG와 ACC의 성능 지표가 우수하게 나왔다. SVM4의 경우에는 고, 중, 저속풍의 비율이 낮기 때문에, Hit 수가 많아지고 Miss의 수가 줄어들어, CSI와 ETS의 성능 지표가 우수하게 나왔다. 강풍 특보에서 중요시되는 성능 지표는 CSI로서, SVM4의 경우 UM에 비해 약 17.0% 우수하고, GP에 비해 49.7% 우수함을 보였다.

표 5 강풍특보 성능 평가 (1)

Table 5 Performance evaluations of windstorm prediction (1)

	UM	GP	SVM1	SVM2	SVM3	SVM4	SVM5
CSI	0.131	0.102	0.077	0.053	0.151	0.153	0.102
PAG	0.160	0.361	0.172	0.496	0.177	0.186	0.231
ACC	0.979	0.992	0.989	0.993	0.979	0.981	0.99
ETS	0.126	0.102	0.074	0.0523	0.145	0.148	0.10

부가적으로 SVM6와 SVM7은 UM 인자 외에 UM 풍속을 함께 사용하여 분류 학습을 수행하였다. 추가 실험한 결과는 표 6과 같다. SVM6가 ACC에서 SVM2와 같게 나온 것을 제외하고는 전반적으로 낮은 성능을 보였다.

표 6 강풍특보 성능 평가 (2)

Table 6 Performance evaluations of windstorm prediction (2)

	SVM6	SVM7
CSI	0.027	0.119
PAG	0.387	0.232
ACC	0.993	0.989
ETS	0.026	0.115

5. 결 론

본 연구에서는 강풍 특보 예측을 위하여 GP 기반의 비선형 보정 방식의 특보 예측 기법과 SVM 기반의 특보 예측 기법을 제주도 지역 101개 격자점에 대해서 5년간의 데이터를 가지고 실험하였다. SVM 학습 데이터의 풍속 비율을 변화시켜 다양한 실험을 수행하였다. 4가지 성능지표가 동시에 만족될 수 없는 다목적 성격이지만, 수치모델인 UM에 비해서 모든 지표가 우수한 SVM 특보 예측 모델을 얻을 수 있었다. 특히, 강풍 특보의 중요 지표인 CSI에 대해서, 제안된 SVM 기법이 UM에 비해 약 17.0%, GP에 비해 49.7% 우수함을 보였다.

풍속 예측을 기반으로 한 GP 보정 방식은 PAG와 ACC 지표에서 우수한 성능을 보였다. 특보 예측에 대한 학습 기반의 접근이 거의 없는 상태에서 상당히 의미있는 결과를 얻었다. 향후 실험을 남한 전지역으로 확대하고, 딥러닝 기법과의 비교도 필요하다고 생각된다.

감사의 글

본 연구는 2015학년도 서경대학교 교내연구비 지원에 의하여 이루어졌음.

References

[1] Korean Meteorological Society, Introduction to Atmospheric Science, Sigma Press, 2009.
 [2] Korea Meteorological Administration, <http://www.kma.go.kr>.
 [3] C. P. Sweeney, P. Lynch, and P. Nolan, "Reducing errors of wind speed forecast by an optimal combination of post-processing methods", Meteorological Applications, Vol. 20, Iss. 1, pp. 32-40. 2013

[4] W. Zhang, J. Wang, J. Wang, Z. Zhao, M. Tian, "Short-term wind speed forecasting based on a hybrid model", Applied Soft Computing, Vol. 13, Iss. 7, pp. 3225-3233. 2013
 [5] H. Wang, G. Wang, G. Li, J. Peng, Y. Liu, "Deep belief network based deterministic and probabilistic wind speed forecasting approach", Applied Energy Vol. 182, pp. 80-93, 2016
 [6] Koza, J. R., Genetic Programming: On the Programming of Computers by Means of Natural Selection, The MIT Press, 1992.
 [7] 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
 [8] 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
 [9] V. Vapnik, Statistical Learning Theory, Wiley, New York, USA, 1998
 [10] United Kingdom Met Office's website : <http://www.metoffice.gov.uk>

저 자 소 개



서 기 성 (Kisung Seo)

1986년 연세대학교 전기공학과 졸업(학사). 1988년 연세대학교 대학원 전기공학과 졸업(석사). 1993년 연세대학교 대학원 전기공학과 졸업(박사). 1999~2003년 Michigan State University, Genetic Algorithms Research and Applications Group, Research Associate, 2002~ 2003년 Michigan State University, Electrical & Computer Engineering, Visiting Assistant Professor, 2011~2012년 Michigan State University, BEACON (Bio/ computational Evolution in Action Consortium) Center, Visiting Scholar, 1993년~현재 서경대학교 전자공학과 교수. 관심분야는 진화연산, 딥러닝, 머신비전, 기상예측, 지능로봇