

Support Vector Regression을 이용한 풍력발전량 예측 시스템 개발

신혜경, 이문환, 이진호
(주) 효성 중공업연구소

Development of the Wind Turbine Power Prediction System Using Support Vector Regression

Hye-Gyeong Shin, Moon-Hwan Lee, Jin-Ho Lee
Hyosung Corporation Power & Industrial Systems R&D Center

Abstract - 신재생에너지는 기후변화협약 및 화석연료의 고갈 등으로 인해 전력계통으로의 도입 필요성은 증가하고 있으나 경제성 부족에 의해 도입 시 많은 제약이 있었다. 그러나 최근 풍력발전기의 경제성이 확보되고 있는 추세이며 일부 유럽 국가를 중심으로 전력계통에 연계하여 운전하고 있다. 특히 스페인의 경우 풍력발전기의 발전량을 예측하는 시스템을 개발하여 풍력발전량의 간헐적인 출력 특성을 보완하고 이용 효율을 향상시킬 수 있도록 다른 발전설비와 연계하여 전력계통을 운영하고 있으며, 풍력발전량을 고려한 예비력을 산정함으로써 경제적이고 안정적인 전력계통을 유지하고 있다. 또한 풍력발전기의 간헐적인 출력 특성을 보완하기 위해 에너지저장장치와의 협조 운영 가능한 시스템을 구축하는 사례가 증가하고 있으며 우리나라의 제주 스마트그리드 실증사업의 Smart Renewable이 이와 같은 경우라 할 수 있다. 본 논문에서는 기계학습이론 중 하나인 SVR을 이용한 풍력발전량 예측 시스템을 개발에 대해 기술하였으며, 행원14호기의 풍력발전량 이력데이터를 이용하여 풍력발전량 예측을 수행하였다.

1. 서 론

일반적으로 풍력발전량 예측은 풍력발전단지의 경제성 평가에 적용되어 왔으며 풍력발전단지의 연간 풍력발전량을 산정하는 것을 의미하였다. 최근 풍력발전기의 경제성이 향상됨에 따라 전력계통에 풍력발전기를 연계하여 운영하려는 시도가 증가하고 있으나 풍력발전기의 간헐적인 출력 특성은 전력계통의 교란을 일으키기 때문에 안정적인 전력계통 운영을 위해서는 대책이 필요하다. 따라서 풍력발전기의 출력 특성을 예측함으로써 교란에 대한 대책을 수립하려는 노력이 증가함에 따라 실시간 풍력발전량을 예측하는 사례가 증가하고 있다. 실시간 풍력발전량을 예측하는 방법은 실시간 풍속을 예측한 후, 예측된 풍속으로 풍력발전량을 산정하는 간접적인 방법과 풍력발전량 이력을 바탕으로 풍력발전량을 예측하는 직접적인 방법이 있으며, Bagging Tree, Multilayer Perceptron, SVM 등과 같은 기계학습이론이 적용되고 있다. 본 논문에서는 직접적인 방법을 이용하여 풍력발전량을 예측하였으며, 예측주기가 10분 이하의 단주기 예측에 적합한 SVR 이론을 적용하였다.

2. 본 론

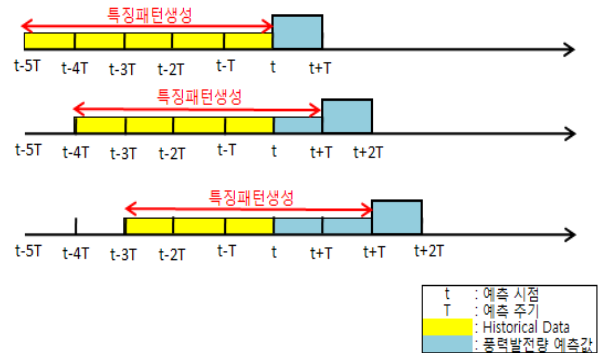
2.1 Support Vector Regression(SVR)의 소개

본 논문에서 제안한 풍력발전량 예측시스템의 기술적 토대가 된 SVR을 소개하고자 한다. SVR은 SVM(Support Vector Machine)을 회귀에 적용하여 개선한 이론이기 때문에 본 절에서는 SVM에 중점 두어 설명하고자 한다. SVM은 유전알고리즘, 신경망 등과 같은 기계학습 이론 중의 하나이며 기계학습이란 임의의 입력과 그것에 대응하는 출력을 반복하여 가르침으로서 특정 입력에 대한 출력을 자동으로 방출시키는 것을 의미한다. 예측대상의 이력데이터를 정보화하여 향후 발생할 데이터를 추론하는 것을 기본으로 한다. 기본적인 분류를 위한 SVM은 입력공간에 Class1과 Class2로 분류할 수 있는 Training Sample이 주어지고 오분류가 최소화 될 수 있도록 분류기준 즉 Maxium-Margin Hyperplane 이라 불리는 선형함수를 도출한다. Training Sample을 분류할 때 앞으로 유입될 sample들을 두 개의 부류로 분류할 수 있는 가능한 무수히 많은 기준 중에서 오분류가 가장 적은 것 같은 분류 기준을 선택하는 것을 의미한다. Hyperlane와 가장 가까운 데이터 즉 가장 분류하기 어려운 데이터들을 Support Vector라고 하며 이 데이터들 간의 간격을 Maxium Margine이라고 한다. Maxium Margine의 간격이 클수록 앞으로 유입될 데이터들의 오분류가 작을 확률이 높으므로 SMR은 Support Vector를 중심으로 Maxium Margine이 최대가 되는 Hyperplane을 도출하는 것이다.

2.2 SVR을 이용한 풍력발전량 예측 시스템의 개요

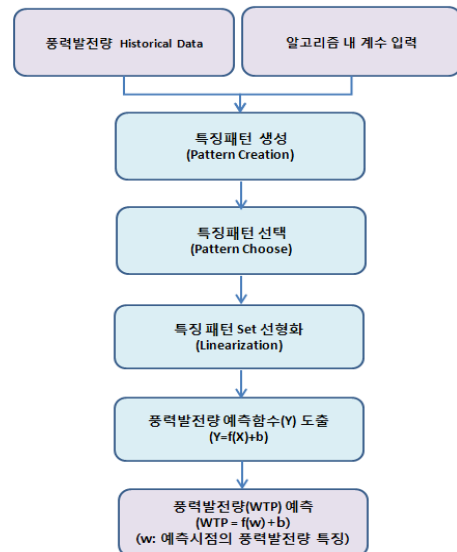
2.2.1 풍력발전량 예측 시스템의 모듈 소개

풍력발전량 예측과 같은 시계열 자료를 예측하는 경우 과거의 패턴이 유지된다는 가정 하에 수집된 데이터의 특징을 분석하고, 추출된 특징 패턴을 바탕으로 연속형 변수의 수학적 모델을 이용하여 회귀분석 혹은 패턴인식을 수행하며 이를 통해 시계열 데이터의 예측값을 도출할 수 있다. 그림 1은 현 시점이 t이고 예측주기가 T인 경우, t, t+T, t+2T 등의 미래 시점의 풍력발전량 예측값을 도출하기 위한 특징패턴 생성 및 예측값 도출의 과정을 도식화 한 것이다.



〈그림 1〉 풍력발전량 예측 과정

본 논문에서 소개하는 풍력발전량 예측 시스템은 예측 대상 풍력발전기의 출력 특징을 추출하여 데이터베이스를 구축하고 예측시점의 풍력발전량의 특징과 유사한 특징패턴들을 이용하여 풍력발전량을 예측한다. 그림 2는 풍력발전량 예측 시스템의 Flow Chart이며 특징패턴 생성, 특징패턴 선택 및 예측 입력데이터의 선형화 단계는 실제 풍력발전량 예측에 앞서 Raw Data의 데이터 정제 단계라고 할 수 있다. 마지막 단계인 풍력발전량 예측 부분에서 SVR을 적용한다.



〈그림 2〉 풍력발전량 예측 시스템의 Flow Chart

특징패턴 생성은 풍력발전기의 이력 데이터를 이용하여 예측 대상 풍력발전기의 출력 특징을 추출하고 정보화하여 데이터베이스를 구축하는 단계이다. 풍력발전기의 출력 특성은 시간대별 풍력발전량 증감 여부에 따른 특징클래스를 나누고 시간대별 변화량(Gradient) 정도에 따라 특징을 정의하였다.

특징패턴 선택은 기 생성된 특징패턴 데이터베이스에서 예측 시점의 풍력발전량 특징과 동일한 특징클래스 선택 후 해당 특징클래스에서 유사한 특징을 선정하여 SVR을 이용하여 예측 함수를 도출할 때 독립변수 및 종속변수로 사용한다.

풍력발전량은 비선형의 데이터로서 선형데이터를 이용하는 SVR에 직접적으로 활용하기 어렵기 때문에 예측 입력데이터 선형화 단계를 거치게 된다. 비선형의 데이터를 선형으로 변환하기 위해 Kernel Function을 이용한다.

풍력발전량 예측함수 도출 단계는 선형화 된 풍력발전량 특징 패턴 Set을 독립변수와 종속변수로 나누어 SVR을 이용하여 해당 풍력발전기에 적합한 예측 함수(F(x))를 도출한다. 도출된 예측함수(F)와 예측시점의 풍력발전량 특징(x)을 이용하여 예측 시점 이후의 풍력발전량을 예측한다.

2.2.2 풍력발전량 예측 시스템의 입력력 데이터

본 논문에서 소개하는 풍력발전량 예측 시스템의 입력데이터는 사용자가 입력하는 계수와 예측 대상 풍력발전기의 Historical Data 로 나뉜다. 출력데이터는 예측시점별 풍력발전량 예측값이며 표1은 입력력데이터를 설명한다.

〈표 1〉 풍력발전량 예측 시스템의 Input/Output Data

구분	적용 단계	계수	정의
입력 데이터	특징패턴 생성	Nsize	풍력발전기의 출력 특징 추출 시 사용하는 샘플링 데이터 수
		Pattern_nMax	풍력발전량의 특징 정의 시 임의의 특징클래스에 저장 가능한 특징패턴의 최대 개수
		m	풍력발전량 특징패턴 생성 시 사용하는 이력 데이터의 개수
	특징패턴 선택	n	예측시점의 특징패턴과 유사한 특징패턴으로 Set을 구성 시 허용 개수
		Tol	유사한 특징패턴을 선정 시 허용 오차
		Predic_N	예측 개수
	예측 입력 데이터 선형화	Kernel Method	비선형데이터인 풍력발전량 데이터를 선형으로 변환하기 위해 적용하는 Kernel Function의 종류
		Sigma	Gaussian Kernel Function의 Width
		npoly	Linear 및 Poly Kernel Function의 다항식 차수
	예측 (SVR)	LossFunction	예측함수의 종속변수와 실측데이터와 오차 계산하여 예측 함수의 품질 정의 시 적용하는 손실함수
		CostF	학습오차를 허용하는 데이터의 개수
		Eta	e-intensive Loss Function의 허용 오차
	입력	Power	예측 대상 풍력발전기의

			이력데이터
출력 데이터	예측결과	WTP	풍력발전량 예측값

2.2.3 사례연구

본 논문에서 소개하는 풍력발전량 예측 시스템의 입력데이터는 13개로 이루어져 있으며 풍력발전량 Historical Data를 제외한 12개의 입력데이터는 사용자가 입력하는 데이터이다. 이러한 12개의 사용자 입력 계수 설정은 현재 연구단계이며 예측 대상에 따라 상이해진다. 따라서 본 논문에서는 행원 14호기(660kW, Vestas)를 대상으로 연구원의 경험을 바탕으로 사용자 입력계수를 설정하였으며, 사용자 입력계수의 변화에 따라 풍력발전량 예측값이 얼마나 상이해 지는 지 살펴보았다.

〈표 2〉 입력계수 변화에 따른 풍력발전량 예측 결과

입력계수	CASE 1	CASE 2	CASE 3	CASE 4	CASE 5	CASE 6
Nsize	5000	5000	5000	5000	5000	
Pattern_nMax	20	29	20	29	20	27
m	5	4	5	4	5	4
n	10	9	10	9	10	8
Tol	1.00E-10	1.00E-10	1.00E-10	1.00E-10	1.00E-10	1.00E-10
Predic_N	500	500	300	300	100	100
Kernel Method	Linear	Linear	Linear	Linear	Linear	Linear
Sigma	5	0.6	5	0.6	5	0.6
npoly	1	1	1	1	1	1
LossFunction	e-insensitive	e-insensitive	e-insensitive	e-insensitive	e-insensitive	e-insensitive
CostF	100	10	100	10	100	10
Eta	0.1	0.5506	0.1	0.5506	0.1	0.5506
결과	CASE1	CASE2	CASE3	CASE4	CASE5	CASE6
오차율(%)	29.51	7.84	26.94	7.44	16.7	7.57

3. 결 론

풍력발전량은 풍속 및 풍력발전기의 기계적 관성 특성에 의해 결정된다. 사례연구에서 보는 것과 같이 동일한 발전기 임에도 불구하고 예측 주기, 적용한 Kernel Function, Loss Function의 종류에 따라 예측 정확도가 상이해 지는 것을 알 수 있다. 따라서 보다 정확한 풍력발전량을 도출하기 위해 예측 대상 설비의 주변요소(다른 풍력발전기와의 간섭정도, 풍향 등) 및 예측 대상 설비의 기계적인 특성을 반영할 수 있는 계수 선정이 선행되어야 하겠다. 본 논문의 예측시스템은 실측데이터를 이용하여 개발되었으나, 실제 풍력발전기의 간헐성을 보완하기 위한 에너지저장장치와 협조 운전에는 반영하지 못하였다. 향후, 실제 풍력발전기와 에너지저장장치의 협조 운전 시 적용함으로써 다양한 운전 상황을 고려할 수 있도록 예측 시스템을 개선할 필요가 있겠다.

[참 고 문 헌]

- [1] Alex J. Smola and Bernhard Scholkopf, "A tutorial on support vector regression", (2004)
- [2] V Cherkassky, "Partial Selection of SVM Parameters and noise estimation for SVM regression", (2004)
- [3] 오일석, "패턴인식", (2008)
- [4] Kusiak, A, Haiyang Zheng, Zhe Song, "Short-Term Prediction of Wind Farm Power: A Data Mining Approach"(2009)