

신재생 에너지 생산량 예측 알고리즘

김지호*

*POSCO ICT

e-mail : kjh532@poscoict.com

요 약

에너지관리 지원 서비스는 공장 내에서 일어나는 전력발전 및 전력할당을 데이터 분석 기법 등을 이용하여 효과적으로 관리하는 것을 목적으로 한다. 특히 그 중에서도 태양광, 풍력 등 친환경 에너지를 이용한 에너지관리 시스템은 비용절감 뿐만 아니라 환경보호 측면에서도 중요한 문제라 할 수 있다. 이들 친환경 에너지를 제대로 이용하기 위해서는 그들의 발전량을 정확히 예측할 필요가 있지만 현재의 시스템에는 가장 기본적인 예측법인 최근접 이웃법을 사용하고 있다. 최근접 이웃법의 경우 노이즈와 아웃라이어에 민감하다는 단점이 있기 때문에 이들 상황에 대처할 수 있는 보다 정교한 예측법이 필요하다.

1. 서론

전통적 관점에서 에너지 관리 시스템은 전문가의 개입이 필요한 수동적인 관리 시스템을 말한다. 이 경우 시스템은 단순히 발전량이나 사용자들의 전력 사용 내역들을 기록하는 정도의 작업을 하기 때문에 이를 이용하여 상황을 모니터링하거나 트렌드를 분석하는 작업은 전문가의 개입이 필요하게 된다. 기존의 시스템에서는 최근접 이웃법을 이용하여 발전량을 예측하도록 에너지 생산량 예측 시스템이 구현되었었다. 장점으로는 구현이 간단하고 계산시간이 빠르다는 장점이 있지만 반면에 노이즈, 아웃라이어에 민감하여 예측량의 변동값 폭이 크다.

2. 기존의 알고리즘

2.1 GPR

비모수적(nonparametric)인 함수 공간 내에서 주어진 데이터와 대상간의 관계를 설명하는 함수를 찾는 방법이다. 여기서 주어진 데이터는 기상 예측 정보를, 대상은 발전량을 의미한다. 최근접 이웃법의 경우 과거의 기상 예측 정보들 중 현재와 가장 비슷했던 때의 발전량을 예측값으로 사용하는 방법으로써 노이즈와 아웃라이어에 민감하다는 단점이 있다. 반면 GPR의 경우 확률 모델에 기반하여 기상 예측 정보와 발전량간의 비선형 관계를 학습하며, 이 모델에 근거하여 현재의 기상 예측 정보가 가질 수 있는 모든 발전량들을 평균하여 예측값으로 사용하기 때문에 최근접 이웃법의 단점을 보완할 수 있다.

2.2 최근접 이웃법

(그림 1)에서 보여지는 것과 같이 과거에 수집된 데이터 중 현재 날씨와 가장 유사한 날의 결과만을 이용한다. 구현이 간단하고 계산시간이 빠르다는 장점이 있지만 노이즈 및 아웃라이어에 민감하다는 단점

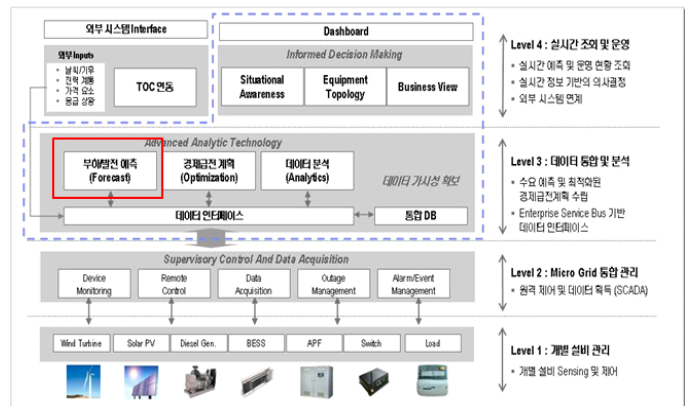
이 있다.

3. 본 논문의 구성

본 논문의 구성은 (그림 2)와 같이 신재생 에너지 관리 시스템 내에서 level 3 데이터 통합 및 분석 단계에서의 부하발전 예측 알고리즘에 대한 설명으로 구성되어 있다. level 2 에서 올라온 데이터들을 데이터 인터페이스를 통해 부하발전 예측을 하고 경제 급전 계획을 하고 데이터 분석을 level 3 단계에서 하게 된다. 그 중에서 부하 발전 예측을 하는데 가장 적합한 알고리즘을 만들어내는 것을 찾아내는 것이 본 논문의 구성이다.



<그림 1> 최근접 이웃법



<그림 2> 본 논문의 전체 구조 시스템

3.1 본 논문의 동작 설명

본 논문의 상세 동작은 아래와 같다.

가) 데이터 학습을 위하여 특정 지역에서 실제로 측정된 풍력 에너지 생산량(단위: KW) 및 풍속기록과 해당 지역에 대한 기상청 관측 정보를 이용하여 (그림3)과 같이 입력데이터로 사용한다. 입력값으로 사용되는 기상 정보 항목에 대한 설명은 (그림4) 와 같다. 예측 알고리즘 모델에 기반하여 학습을 하기 위해서는 데이터를 군집화하는 과정이 먼저 이루어지게 된다. 군집화 알고리즘 중 가장 널리 사용되는 k-mean clustering 알고리즘을 사용한다.

나) 원활한 예측을 위해 전체 데이터 중에서 일부는 다음의 이유들로 인해 수정 및 제거한다. 먼저 기상청 관측 정보들에서는 일부 기록이 누락된 경우가 종종 있었는데 이들을 상황에 따라 아예 사용하지 않거나 전후의 측정값들을 이용하여 평균을 취하는 방법을 이용하여 수정한다. 예를 들어 측정값들에서는 측정된 발전량이 설치된 풍력 발전기의 발전 한계량인 17kw를 초과하는 경우, 풍속이 0m/s인데 발전량이 0보다 큰 경우, 풍속이 6m/s 이상인데 발전량이 0 으로 기록되는 경우에는 정상적인 측정이 이루어지지 않았다고 판단하여 데이터에서 제외하게 된다.

다) 과거의 학습데이터를 통해 HGPR의 알고리즘이 다른 알고리즘보다 정확성이 높다는 것을 알 수 있었다. 과거의 기상정보와 과거패턴을 통해 미래의 예측 생산량을 (그림5)와 같이 얻을 수가 있다.(그림6)에 나와 있듯이 미래의 예측 에너지 생산량을 얻기 위해서는 날짜, 시간, 풍속, 강수량, 풍향을 입력하면 (그림7)과 같은 예측 결과값을 얻을 수 있게 된다.

예측 결과값을 얻기 위한 평균 계산은

$$\bar{f}(x_*) = \mathbf{k}_j^\top (\boldsymbol{\alpha}_j - [\tilde{\boldsymbol{\mu}}]_j \mathbf{d}_j) + ([\mathbf{K}_g]_{:,j})^\top (\boldsymbol{\mu} - \Delta \tilde{\boldsymbol{\mu}})$$

이다.

위의 값들을 얻기 위한 과정은 아래와 같다.

- 목적 함수(marginal likelihood)

$$\begin{aligned} \mathcal{L} &= \frac{1}{2} \log |\boldsymbol{\Sigma}_N| + \frac{1}{2} \mathbf{y}^\top \boldsymbol{\Sigma}_N^{-1} \mathbf{y} \\ &= \frac{1}{2} \left(\log |\tilde{\mathbf{D}}| + \log |\mathbf{K}_g| + \log |\mathbf{A}| + \mathbf{y}^\top \tilde{\mathbf{D}}^{-1} \mathbf{y} - \mathbf{y}^\top \tilde{\mathbf{D}}^{-1} \mathbf{H} \mathbf{A}^{-1} \mathbf{H}^\top \tilde{\mathbf{D}}^{-1} \mathbf{y} \right) \\ &= \frac{1}{2} \left(\sum_{j=1}^Q \log |\tilde{\mathbf{D}}_j| + \log |\mathbf{K}_g| + \log |\mathbf{A}| + \sum_{j=1}^Q \boldsymbol{\alpha}_j^\top \mathbf{y}_j - \boldsymbol{\mu}^\top \tilde{\boldsymbol{\mu}} \right), \end{aligned}$$

- 목적 함수의 1차 미분

각 군집의 센터들간의 kernel 함수를 계산하는데 필요한 파라미터에 대한 미분:

$$\mathbf{M} = \mathbf{A}^{-1} + \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^\top, \quad \frac{\partial \mathcal{L}}{\partial \phi} = \frac{1}{2} \text{tr} \left\{ \left(\mathbf{K}_g^{-1} - \mathbf{K}_g^{-1} \mathbf{M} \mathbf{K}_g^{-1} \right)^\top \frac{\partial \mathbf{K}_g}{\partial \phi} \right\}$$

각 군집 안 데이터들간의 kernel 함수를 계산하는데 필요한 파라미터, noise variance parameter에 대한 미분:

$$\frac{\partial \mathcal{L}}{\partial \phi} = \frac{1}{2} \sum_{j=1}^Q \text{tr} \left\{ \boldsymbol{\Gamma}_j^\top \frac{\partial \tilde{\mathbf{D}}_j}{\partial \phi} \right\}$$

$$\boldsymbol{\Gamma}_j = \tilde{\mathbf{D}}_j^{-1} - \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j^\top + \mathbf{d}_j \left(2[\tilde{\boldsymbol{\mu}}]_j \boldsymbol{\alpha}_j^\top - [\mathbf{M}]_{jj} \mathbf{d}_j^\top \right)$$

$$\boldsymbol{\alpha}_j = \mathbf{D}_j^{-1} \mathbf{y}_j, \quad [\mathbf{d}_j]_l = \sum_{l=1}^{N_j} [\mathbf{D}_j^{-1}]_{l,l}$$

4. 알고리즘 간에 비교

예측 방법 간 비교에서는 HGPR 를 다른 세가지 예측 방법들(최근접 이웃법, support vector regression(SVR), Gaussian process regression(GPR)) 과 비교한다.

여기서는 총 세 개의 예측 방법들과 성능을 비교하였다. 첫 번째로 비교한 방법은 최근접 이웃법으로써, 현재 에너지 관리 시스템에 근접 이웃 수 1의 설정으로 구현 돼 있다. 그 외에 regression 기법으로써 많이 사용되는 SVR 과 HGPR 의 원 모델인 GPR 과도 성능을 비교하였다. (그림 8), (그림 9), (그림 10), (그림 11), (그림 12)의 결과들에 따르면 최근접 이웃법의 경우 근접 이웃 수에 관계없이 항상 HGPR 보다 예측 성능이 좋지 않았고, SVR 의 경우도 역시 cost 파라미터에 상관없이 HGPR 보다 성능이 좋지 않았다. GPR 의 경우 같은 수의 학습데이터에서는 평균적으로 0.2 정도 예측 성능이 좋았지만 10 분마다 학습을 다시 해야 하는 시스템의 특성을 고려하면 적합하지 않다는 결론을 얻었다. (GPR 의 경우 학습데이터가 약 2,000 개일 때 10 분 가까이 계산 시간이 필요하지만 이 경우의 MAE 는 약 1.93으로써 더 적은 시간에 1.5 이하의 MAE 를 얻을 수 있는 HGPR 보다 떨어짐)

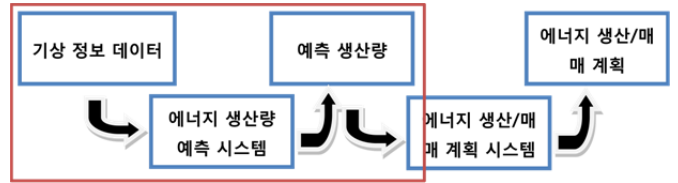
5. 알고리즘 설명

5.1 HGPR 알고리즘

HGPR 은 데이터들이 일반적으로 군집화하는 경향이 있다는 점을 이용하여 GPR 의 시간 복잡도를 줄이면서 군집에 따라 다를 수 있는 데이터와 대상간의 관계를 학습 모델에 반영하기 위해 제안된 알고리즘이다. HGPR 을 설명하는 확률 모델은 (그림 13)과 같다. 여기서 x 는 데이터(기상 예측 정보), y 가 대상(발전량)에 해당하며 c 는 각 군집들의 센터, f 는 데이터와 대상간의 비선형 관계를 표현하는 함수, g 는 각 군집들의 센터들간의 관계를 표현하는 함수를 나타낸다. 여기서 f 와 g 는 사전에 알 수 없는 변수들이기 때문에 HGPR 학습을 통해서 구한다. 노이즈 및 아웃라이어에 크게 영향을 받지 않는다는 장점은 있다. 하지만 계산시간이 느리다는 단점이 있지만 파라미터를 조정하여 계산시간이 5 분 이내가 되도록 구현해서 보완할 수가 있다.

5.2 기타 고려사항: 데이터 군집화

HGPR 확률 모델에 기반하여 학습을 하기 위해서는 데이터들을 군집화하는 과정이 먼저 이루어져야 한다. (모델에 이미 데이터들(x)이 군집을 이루고 있고, 각 군집의 센터들도 사전에 주어지도록 설계가 되어있기 때문) 본 프로그램이 예측하고 하는 풍력 에너지 발전량은 어느 정도 시간이나 계절의 영향을 받기는 하지만 어떤 식으로 군집화해야 하는 지 명확한 경계를 사람이 일일이 정하기는 어렵기 때문에 자동 군집화 알고리즘을 사용한다. 여기서는 군집화 알고리즘 중 가장 널리 사용되는 k-mean clustering 알고리즘을 사용한다.



<그림 5> 신재생 에너지 생산량 예측 기술 개발

날짜	시간	풍속	강수량	풍향	기온
2012-06-03	2:00:00	5.94	48	90.4	15
2012-06-03	1:50:00	5.05	48	77.1	16
2012-06-03	1:40:00	4.65	49	85	12
2012-06-03	1:30:00	3.88	50	87.5	17
2012-06-03	1:20:00	5.72	48	90	15
2012-06-03	1:10:00	7.12	50.5	78	19
2012-06-03	1:00:00	7.60	51	79	14
2012-06-03	0:50:00	5.89	47	95	15
2012-06-03	0:40:00	7.71	49	70	16

<그림 6> 미래 생산량 예측을 위한 input data

날짜	시각	예측생산량(kw)
2010.6.03	2:00:00	9.255185767
2010.6.03	1:50:00	9.259659036
2010.6.03	1:40:00	9.990810723
2010.6.03	1:30:00	9.396060425
2010.6.03	1:20:00	6.237619645
2010.6.03	1:10:00	8.289291089
2010.6.03	1:00:00	7.75512145
2010.6.03	0:50:00	5.9382281
2010.6.03	0:40:00	6.328490917

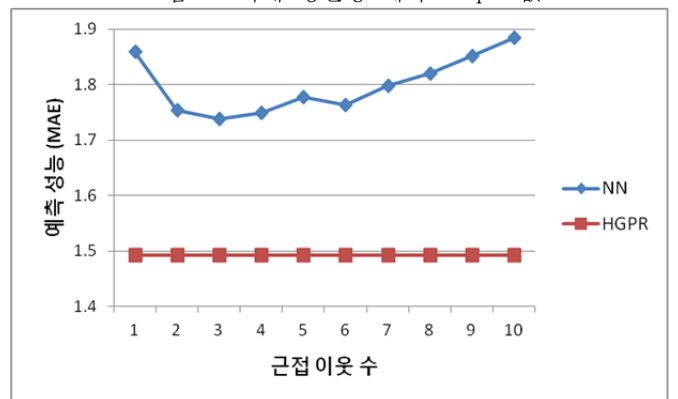
<그림 7> 미래 생산량 예측 output 값

Wind Forecast Sample Data																		
		Wind Speed At 70m																
3	Date	Time	85m	70m	50m	Output (kw)	사분	강수15	강수60	강수12H	일강수	기온	풍향1	풍속1	풍향10	풍속10		
4	2010-04-20	2:10:00	5.93	0.00	0.21	10.97	210	0	0	38.5	0.5	15	76.9	ENE	3.9	75.4	ENE	4.3
5	2010-04-20	2:00:00	5.31	0.00	0.21	9.84	200	0	0	40	0.5	15	77.9	ENE	3.7	73.2	ENE	3.7
6	2010-04-20	1:50:00	4.79	0.00	0.21	9.74	150	0	0	41.5	0.5	15.1	75.7	ENE	3.4	74.2	ENE	3.8
7	2010-04-20	1:40:00	5.94	0.00	0.21	8.34	140	0	0	42.5	0.5	15.2	71.4	ENE	4.5	80	E	3.7
8	2010-04-20	1:30:00	5.05	0.00	0.21	8.58	130	0	0.5	44	0.5	15.2	73.2	ENE	3.4	76.6	ENE	3.7
9	2010-04-20	1:20:00	4.65	0.00	0.21	6.80	120	0	0.5	45.5	0.5	15.2	76.6	ENE	4.1	75.3	ENE	3.9
10	2010-04-20	1:10:00	3.88	0.00	0.21	5.28	110	0	0.5	46.5	0.5	15.2	70.7	ENE	3.1	74.1	ENE	3.2

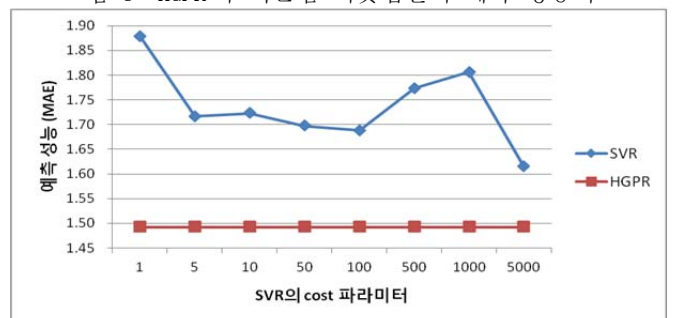
<그림 3> 입력 데이터

기상 정보 항목 설명			
순번	항목명	설명	단위
1	WS85	해발 85m 에서의 풍속	m/s
2	WS70	해발 70m 에서의 풍속	m/s
3	WS50	해발 50m 에서의 풍속	m/s
4	PRC15	15분간 강수량	mm
5	PRC60	60분간 강수량	mm
6	PRC12	12시간 강수량	mm
7	PRC24	일강수량	mm
8	TMP	기온	℃
9	WS1M	1분간 평균 풍속	m/s
10	WS10M	10분간 평균 풍속	m/s
11	WD1M	1분간 풍향 변화	°
12	WD10M	10분간 풍향 변화	°

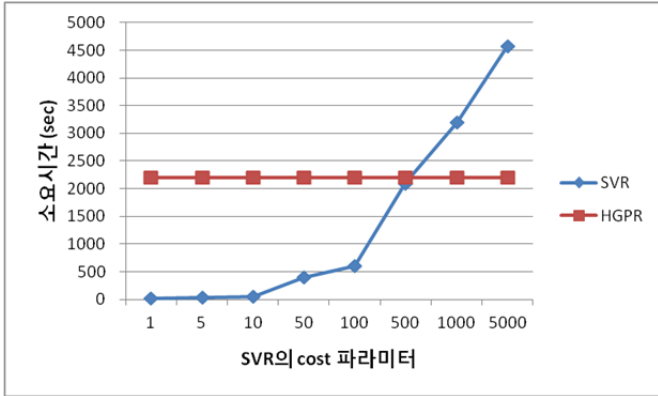
<그림 4> 기상 정보 항목 설명



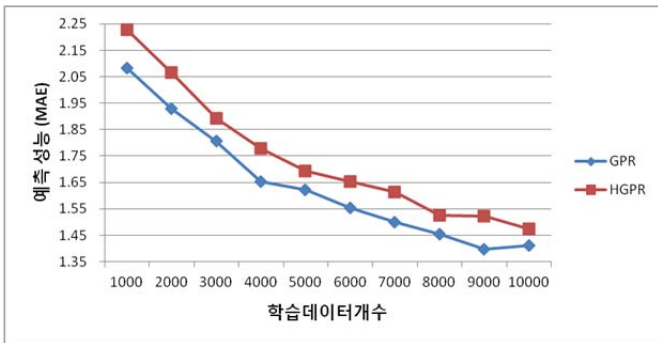
<그림 8> HGPR 과 최근점 이웃법간의 예측 성능비교



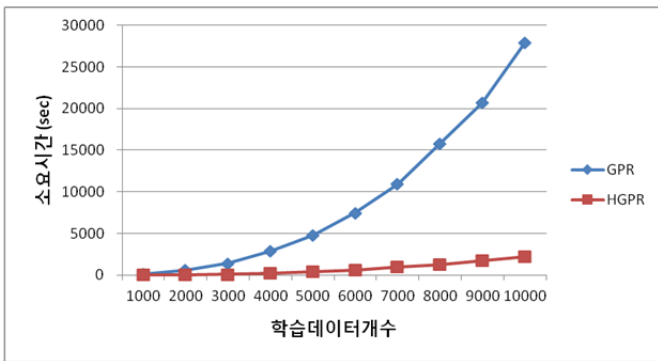
<그림 9> HGPR 과 SVR 간의 예측 성능비교



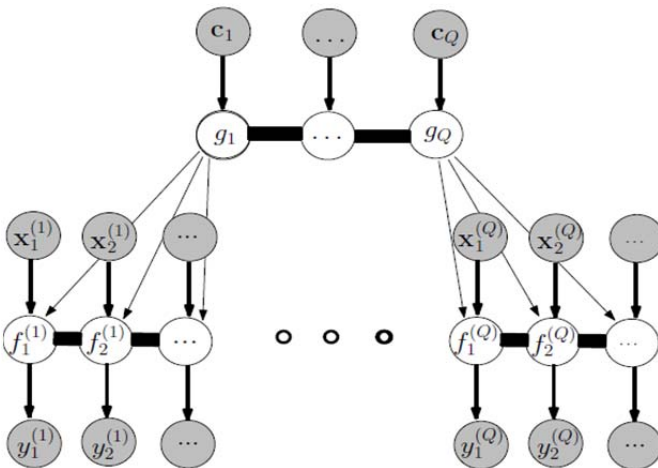
<그림 10> HGPR 과 SVR 간의 소요시간 비교



<그림 11> HGPR 과 GPR 간의 예측 성능 비교



<그림 12> HGPR 과 GPR 간의 예측 성능비교



<그림 13> HGPR 모델

3. 연구 결론 및 토의

본 연구로 인해 전통적 에너지 관리 시스템과 가장 크게 달라지는 점은 전문가의 판단을 기다리는 것이 아니라 시스템이 자체적으로 정해진 모델에 따라 데이터베이스에 저장된 기록들을 분석하고 따라서는 판단까지 내리게 된다. 또한 이상이 발생하였을 때 원인을 쉽게 찾을 수 있도록 특정 설비에 대해 영향을 주는 인자들에 대해 정의한 뒤, 이들에 대해 민감도 분석, 추세 분석, 연관성 검증, 인과관계 도출의 과정을 거쳐서 어떤 인자들이 해당 설비의 에너지 효율에 주된 영향을 미치는 지에 대해서도 분석하게 된다. 마지막으로 HGPR 가 최근접 이웃법 알고리즘보다는 전반적으로 예측이 크게 빗나가는 경우가 적게 발생하고, GPR 보다는 학습데이터가 늘어나면 늘어날수록 학습시간이 현저히 줄어들게 된다. 또한 <그림 8> ~ <그림 11> 을 통해 알 수 있듯이 기존의 방법과는 다르게 HGPR 는 외부 요인에 의해 불필요한 영향을 받지 않는다는 것을 알 수가 있다. 이런 부분을 통해 기존의 방법들이 가지고 있던 노이즈나 아웃라이어에 크게 영향을 받던 문제점들을 보완된다는 사실을 알 수 있다.

참고문헌

- [1] BIR RESEARCH GROUP, 2012 년 신재생에너지산업과 풍력사업 동향, 비아이알, 2012
- [2] 최대연간에너지 생산량을 위한 MADS 기반의 풍력발전기 최적 설계. 2008
- [3] 데이코산업연구소, 태양광사업 기술 시장의 실태와 전망 2009, 산업교육연구소, 2009
- [4] 제프시겔, 재생 가능 에너지에 투자하라, 휴먼앤북스, 2009
- [5] 윤재현, 지금 당장 원자재 공부 시작하라, 한빛비즈, 2009