

앙상블 기법을 이용한 태양광 발전소 고장 예측 개선

*장문중, **나익채, #김영훈

한양대학교 ERICA, 컴퓨터공학과

*seung10521@naver.com, {**njoy, #nongaussian}@hanyang.ac.kr

Improvement of Predicting Failure Rate of Photovoltaic System using Ensemble Methods

*Munjong Jang, **Ickchae Na, #Younghoon Kim

Hanyang University ERICA, Dept. of Computer Science

요 약

최근 태양광 발전사업의 투자 수요가 증가하고 있으며, 이에 따른 태양광 발전시스템(PV시스템)의 신뢰성 및 발전 효율 향상 등을 확보할 수 있는 모니터링 시스템의 중요성이 부각되고 있다. 본 논문에서는 데이터를 앙상블 기법으로 분석하여 알려진 자동 분류 기법과 앙상블 기법을 비교해보고, 이를 바탕으로 PV시스템 고장 예측의 정확도를 향상 시키고자 한다.

1. 서 론

태양광 발전시스템의 발전단가가 그리드패리티¹에 도달하였고, 2018년 이내에 중국을 제외한 모든 나라가 그리드패리티를 달성할 것으로 예상되고 있다. 그리드패리티의 확보와 마이너스금리 시대가 도래함에 따른 태양광 발전사업의 투자 수요가 증가되고 있다. 이에 따른 PV시스템의 신뢰성 및 발전 효율 향상 등을 확보할 수 있는 모니터링 시스템의 중요성이 부각되고 있다.

효율적인 통합관리 시스템의 개발을 통해 우리는 PV 시스템의 유휴 시간을 최소화함으로써 발전생산량을 최대화하고, 고장 방지 및 신속한 고장 복구, 고장 예측 및 발전량 예보를 통한 구성 부품 및 인력 관리의 최적화와 계획된 일정에 따른 효율적인 유지보수로 시스템의 안정성과 신뢰성을 최대화하고자 한다. 본 논문에서는 태양광 발전 시스템의 고장 데이터를 알려진 자동분류 기법을 앙상블시켜 단일 기법을 사용하는 것 보다 비교적 안정적인 예측이 가능함을 확인하였다.

2. 관련 연구

앙상블 기법에 관한 연구들 중 [1]에서는 오차율 최소화를 위해서 Stacking 기법을 도입하였고, [2]에서는 문자 인식에 사용되는 Boosting 알고리즘의 성능에 관한 연구가 진행되었다. [3], [4]에서는 각각 Bagging과 Voting 기법을 고안하여 예측의 정확도를 향상시키는 방법을 제안하고 있다. 본 논문에 잘 서는 상술한 앙상블 기법들을 이용하여 발전시스템의 고장 예측 정확도를 향상 시키고자 한다.

3. 분석 방법

주어진 발전소 발전량 데이터는 두 발전소에서 각각 수집된 일사량, 풍속, 기온, AC/DC발전량, AC/DC전압, AC/DC전류, 고장 여부이며 각 발전소에서 수집된 데이터를 PV1과 PV2으로 표기한다. 고장시의 수집데이터와 정상 작동 시의 수집데이터는 각각 수리 시점 이전 3일과 수리 이후 7일의 데이터를 추출하여 사용하였다. 또한 AC와 DC전력값이 0인, 발전이

¹ 화석연료 발전단가와 신재생에너지 발전단가가 같아지는 시기를 말한다.

<표 1> PV별 속성값의 종류

PV시스템	PV1	PV2
속성	Insolation : 일사량	
	atTemp : 기온	
	windSpeed : 풍속	
	DCV, ACV : 직류/교류 전압	
	DCI, ACI : 직류/교류 전류	
	DCP, ACP : 직류/교류 발전량	
	Broken : 고장 여부	
	X	
	sTemp1 :표면온도1	
	sTemp2 :표면온도2	

이루어지지 않는 시간대의 데이터는 제거하였다. PV2에는 PV1에는 없는 두개의 온도 센서에서 측정된 표면 온도 속성이 더 있어 <표 1>과 같이 각각 10차원과 12차원의 속성값들로 구성되어있다.

본 연구에서는 데이터에 k-NN(k-Nearest Neighbors)[5], 신경망(Neural Network)[6], 로지스틱 회귀(Logistic Regression)[7], SGD(Stochastic Gradient Descent)[8]등의 분류기법들에 앙상블 기법인 Boosting, Bagging, Stacking, Voting등을 적용하여 태양광 발전시스템의 고장 예측 정확도를 향상시키고자 한다.

- Boosting[2]: Boosting은 자동분류기법에 의해 잘못 분류된 학습데이터에 대해 이 들의 분류정확도를 높이기 위해 재차 집중적인 분류를 수행하여 새로운 분류 규칙을 만드는 단계를 반복하는 방법이다.
- Bagging[3]: Bagging은 주어진 데이터에 대해서 랜덤샘플링을 통해 얻은 여러 개의 부트스트랩(bootstrap) 자료를 생성하고 각 부트스트랩 자료를 모델링 한 후 결합하여 최종의 예측 모형을 산출하는 방법이다.
- Stacking[1]: Stacking은 메타 분류기와 일반 분류기를 선택해 일반 분류기의 결과값을 메타 분류기를 이용해 다시 분석한다.
- Voting[4]: 서로 다른 모델에서 예측을 조합 또는 같은 형태의 모델에서의 예측 후 평균화를 위해 voting을 한다.

4. 실험

실험 환경은 Intel i5-6500 3.20GHz CPU, 8GB RAM,

Window 7 OS에서 수행했으며 WEKA 3.8 Version을 이용하였다. PV시스템1(PV1), PV시스템2(PV2)에 대해 k-NN, 신경망, 로지스틱 회귀, SGD와 같은 분류 기법들을 사용하여 태양광 발전 시스템의 고장을 예측하였고, 고장이라고 예측된 발전소들 중 실제로 고장난 발전소의 비율인 정밀도(Precision; PRC)와 고장난 발전소들 중 실제로 예측된 발전소의 비율을 나타내는 재현율(Recall; RCL), 그리고 정밀도와 재현율의 조화평균인 F1-Score(F1)를 측정하였다. 또한 앙상블 기법인 Boosting, Bagging, Stacking, Voting를 차례대로 적용하였으며 각 기법들에 대하여 정밀도, 재현율, F1-Score를 측정하였다. Boosting과 Bagging기법은 k-NN, 로지스틱 회귀, 신경망, SGD를 적용해 산출된 값들의 평균값을 구했다. Stacking기법은 상술한 4가지 분류 기법 중 하나를 메타 분류기로 정할 때마다 나머지 3가지 기법들 각각이 일반 분류기가 되는 경우가 발생하기 때문에, P(4,2)가지의 경우에 대한 평균값을 구하였다. Voting 기법은 4가지 분류 기법들 중 2가지씩을 조합해 $\binom{4}{2}$ 가지의 경우에 대한 평균값을 구하였다. 각 측정값들은 소수 둘째자리까지 반올림해 각각 <표 2>, <표 3>에 나타내었다.

<표 2> PV별 정밀도, 재현율, F1-Score

기법	PV1			PV2		
	PRC	RCL	F1	PRC	RCL	F1
k-NN	0.67	0.70	0.69	0.72	0.70	0.71
로지스틱 회귀	0.97	0.48	0.64	0.83	0.75	0.79
신경망	0.89	0.69	0.77	0.87	0.83	0.85
SGD	0.93	0.42	0.58	0.83	0.75	0.79
평균	0.87	0.57	0.67	0.81	0.76	0.79

기법	PV1			PV2		
	PRC	RCL	F1	PRC	RCL	F1
Boosting	0.79	0.60	0.67	0.82	0.79	0.80
Bagging	0.87	0.55	0.66	0.80	0.76	0.78
Stacking	0.80	0.57	0.65	0.77	0.71	0.74
Voting	0.84	0.55	0.64	0.82	0.75	0.78
평균	0.83	0.57	0.66	0.80	0.75	0.78

<표 3> 앙상블 기법의 PV별 정밀도, 재현율, F1-Score

앙상블 기법이 단일 기법보다 일관적인 수치를 보이며 특히 재현율의 편차가 크지 않아 단일 기법만을 사용할 때보다 안정적인 예측이 가능함을 확인할 수 있다.

5. 결과

본 연구에서는 k-NN, 신경망, 로지스틱 회귀, SGD 등의 분류기법들을 각각 적용하여 PV시스템의 고장을 예측해 보았고, Boosting, Bagging, Stacking, Voting 등의 앙상블 기법으로 상술한 분류 기법들을 조합하여 PV시스템 고장을 다시 한번 예측해보았다. 또한 앙상블 기법을 사용할 때 단일 기법을 사용할 때보다 안정적인 예측이 가능함을 확인할 수 있었다. 향후 연구 과제는 앙상블 기법을 개선해 정밀도와 재현율 그리고 F1-Score를 향상시키고, 이를 바탕으로 PV시스템의 고장을 더욱 정확하게 예측하는 것이다.

사사의 글

본 연구는 산업통산자원부 에너지기술개발사업(20153010011980, 과제명: 태양 광발전 운영효율 향상을 위한 통합관리 시스템 개발)의 연구비 지원으로 수행

참고 문헌

[1] D. Wolpert, "Stacked Generalization" *Neural Networks*, vol. 5, no. 2, pp. 241-259, 1992.
 [2] H. Drucker, R. Schapire and P. Simard, "Boosting Performance in Neural Networks" *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 4, pp. 705-719, 1993.
 [3] L. Breiman, "Bagging predictors" *Machine learning*, vol. 24, no. 2, pp. 123-140, 1996.
 [4] R. Schapire, Y. Freund, P. Bartlett and W. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods" *Annals of statistics*, pp. 1651-1686, 1998.
 [5] K. Fukunaga and P. Narendra, "A branch and bound algorithm for computing k-nearest neighbors" *IEEE*

transactions on computers, vol. 100, no. 7, pp. 750-753, 1975.

[6] L. Hansen and P. Salamon, "Neural network ensembles" *IEEE transactions on pattern analysis and machine intelligence*, vol. 12, pp. 993-1001, 1990.

[7] D. Jennings, "Judging inference adequacy in logistic regression" *Journal of the American Statistical Association*, vol. 81, no. 394, pp. 471-476, 1986.

[8] W. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique" *Signal Processing*, vol. 6, no. 2, pp. 113-133, 1984.