

Técnicas de aprendizaje computacional aplicadas a la fotovoltaica

Lucía Serrano-Luján, José Manuel Cadenas*, Antonio Urbina

Departamento de Electrónica, Tecnología de computadores y Proyectos. ETSI Telecomunicaciones.
Campus Muralla del Mar 30202 Cartagena (Murcia).

* Departamento de Ingeniería de la Información y las Comunicaciones, Facultad de Informática, Universidad de Murcia

E-mail: lucia.serrano@upct.es

Resumen. En este artículo mostramos los resultados de la aplicación de varios métodos de aprendizaje computacional sobre datos procedentes de una instalación fotovoltaica de 222kWp de CdTe. Los datos de uno de los 30 grupos de módulos han sido estudiados en profundidad para predecir el estado en que se encuentra cuando el rendimiento es bajo: afectado por sombra fija, por sombra variable, los módulos están sucios, error ajeno a dichas circunstancias, etc. o bien identificar que está generando electricidad de forma adecuada.

1. Introducción

La atención hacia la investigación sobre la fiabilidad de los sistemas fotovoltaicos está incrementando debido al rápido crecimiento de instalaciones fotovoltaicas en zonas residenciales, edificios comerciales, así como en bases militares [1]. El principal interés de los investigadores en la energía solar hace referencia al diseño y optimización de esta tecnología, así como mejorar su eficiencia [2]. Son muchos los estudios que hacen referencia a la evaluación de la fiabilidad y el riesgo en sistemas fotovoltaicos a gran escala como un medio de conseguir beneficios, tanto para las compañías como para los clientes.

Hasta la fecha son muchos los algoritmos de aprendizaje computacional que han sido aplicados a energías renovables. Algunos autores como [3] y [4] aplican algoritmos *fuzzy* de toma de decisiones, basado en conocimiento experto, para decidir cuándo conectar esta tecnología a la red y cuándo almacenar su energía, otros lo aplican para extender en el tiempo la operatividad de una bomba de agua.

Normalmente, cuando se aplican estas técnicas de aprendizaje en la medida del rendimiento de las tecnologías fotovoltaicas, se aplican de cara a estimar la producción energética, y en concreto en métodos de predicción de condiciones atmosféricas.

Este estudio se centra en describir una metodología que pueda controlar el rendimiento de un grupo de módulos de una instalación fotovoltaica a gran escala, identificando si es posible cuándo ocurre un error y cuál es la causa que lo produce. Principalmente trata de obtener un algoritmo que controle el rendimiento de un grupo de módulos. En un futuro estudio se pretenderá además prescindir de las condiciones medioambientales, ya que no todas las instalaciones poseen sensores ambientales.

Al aplicar técnicas de manejo de datos y algoritmos de aprendizaje computacional a bases de datos de una

instalación fotovoltaica una amplia variedad de relaciones entre los parámetros pueden encontrarse.

2. Al encuentro de relaciones entre parámetros fotovoltaicos

El manejo de datos es un área de la ciencia de la computación donde las herramientas informáticas son desarrolladas para encontrar relaciones en grandes grupos de datos. Los métodos de aprendizaje computacional es un sub-área del manejo de datos que se centra en utilizar técnicas para encontrar y describir patrones en los datos como una herramienta para explicar los datos y realizar predicciones [5].

Este estudio se basa en los datos que han sido almacenados los últimos años procedentes de un generador fotovoltaico de 222 kWp de potencia instalada, con módulos de tipo CdTe. Múltiples datos como la energía generada, la diferencia de potencial, etc. han sido almacenados cada 5 minutos durante unos 3 años. A día de hoy la base de datos contiene más de 200,000 registros (con 16 atributos) para cada uno de los 30 inversores que forman parte del generador.

Este generador está conectado a red y se considera de larga escala. Las sombras fijas, sombras variables, malas condiciones atmosféricas o módulos sucios son circunstancias frecuentes que pueden provocar un descenso en el rendimiento.

La Fig. 1 muestra el rendimiento de cuatro grupos de módulos a lo largo de un mes, comparado con el rendimiento medio del generador. Uno de los grupos manifiesta un rendimiento inferior al medio debido a que sus módulos son afectados por una sombra fija a ciertas horas de la mañana.

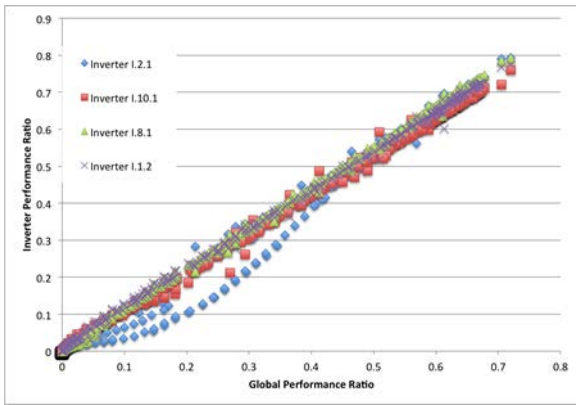


Figura 1. Comparación entre el rendimiento de un generador fotovoltaico de gran escala con el rendimiento de cuatro de los grupos de módulos que lo componen.

3. Aplicando técnicas de aprendizaje computacional. Metodología.

3.1. Pre-procesamiento de los datos

Una de las partes fundamentales de este estudio, y la que más tiempo requiso, fue el pre-procesamiento de los datos. Algunos de los criterios que se utilizaron para seleccionar los registros adecuados fueron: datos enviados sólo durante horas de sol, que no pierdan ningún campo, el estado del inversor debe ser distinto al estado “error” (0).

De los 30 inversores, el estudio se centró en un único inversor. 30,550 registros cumplieron los criterios.

Además de los campos que se envían automáticamente desde el generador fotovoltaico, se ha añadido un atributo “rendimiento” o “performance ratio” (PR), definido por la siguiente ecuación:

$$PR = \frac{Pac}{k \cdot \frac{G}{G_{STC}}}$$

Donde:

Pac: potencia generada (kW/m²)

G_{STC}: Irradiación en condiciones estándar (1000 W/m²)

G: Irradiancia sobre el generador

k: potencia pico del grupo de módulos

3.2. Aplicación del conocimiento experto

Cada uno de los registros han sido etiquetados por un experto siguiendo el siguiente formato de etiquetas:

3.3 Aplicando algoritmos

Los algoritmos aplicados son aquellos identificados como útiles para nuestro propósito: detectar y clasificar estados del grupo de módulos.

Tabla 1. Etiquetas que identifican los posibles estados del grupo de módulos fotovoltaicos.

Nombre	Descripción
F	Sombra fija grande (Potencia = 0)
F1	Sombra fija pequeña (Potencia > 0)
T	Nubes o lluvia (Potencia = 0)
T1	Nubes o lluvia (Potencia > 0)
M	Módulos sucios
OK	Rendimiento adecuado
S	Sombra variable (Potencia > 0)
E1S1	Posible error o sombra variable grande (Potencia = 0)
E2S2	Posible error o sombra pequeña variable (Potencia > 0)

Tabla 2. Algoritmos de aprendizaje aplicados al problema

Método de aprendizaje	Descripción
	<i>Clasificador modelado...</i>
C4.5	<i>...en forma de árbol de decisión</i>
Random forest	<i>...mediante un ensamble de árboles de decisión</i>
Ripple-Down Rule learner	<i>...mediante reglas por defecto y sus excepciones</i>
PART	<i>...mediante un conjunto de reglas tipo PART</i>
Reglas conjuntivas	<i>...mediante una regla tipo conjuntiva</i>

3. Resultados

Partiendo de las 30,550 instancias con 35 atributos que fueron etiquetadas por un experto, la siguiente tabla muestra los resultados de analizar los resultados durante el proceso de aprendizaje en términos de precisión de la predicción. Se ha utilizado validación cruzada de 3x5.

Tabla 3. Resultados de aplicar los métodos de aprendizaje. Precisión media y error cuadrático medio.

Interpretable	Método	μ	Σ
✓	C4.5	99.97	0.009
✗	Random Forest	99.96	0.015
✓	Ripple-Down Rule	99.95	0.012
✓	PART	99.95	0.012
✓	Conjuntive Rule	99.61	0.113

Tras analizar estos algoritmos se ha seleccionado el C4.5, puesto que es interpretable y tiene la mayor precisión en la predicción. En concreto se ha obtenido un árbol de 25 hojas y un tamaño de 749, como muestra la Fig. 2.

```

PR <= 0.699
| Upv_Soll <= 786.5
| | PR <= 0.499
| | | Gd <= 28
| | | | Pac4 <= 86.33
| | | | | PR1 <= 0.498
| | | | | | PR3 <= 0.493
| | | | | | | PR2 <= 0.499
| | | | | | | | PR5 <= 0.496: M (212.0/1.0)
| | | | | | | | PR5 > 0.496: S (14.0)
| | | | | | | | PR2 > 0.499: S (17.0)
| | | | | | | | PR3 > 0.493: S (25.0)
| | | | | | | | PR1 > 0.498: S (40.0)
| | | | | Pac4 > 86.33
| | | | | | G <= 14: S (291.0/2.0)
| | | | | | G > 14
| | | | | | | Ipv <= 0.14: T1 (5.0)
| | | | | | | Ipv > 0.14: S (30.0/2.0)
| | | | Gd > 28
| | | | | Ipv <= 0.34
| | | | | | Gd <= 30
| | | | | | | Sol <= 23.45: T1 (13.0)
| | | | | | | Sol > 23.45: S (8.0/1.0)
| | | | | | | Gd > 30: T1 (1001.0)
| | | | | Ipv > 0.34
| | | | | | TIMESTAMP hora <= 49800
| | | | | | | Iac Ist <= 5.59: T1 (4.0)
| | | | | | | Iac Ist > 5.59: S (3.0)
| | | | | | | TIMESTAMP hora > 49800: S (15.0)
| | PR > 0.499
| | | PR5 <= 0.698
| | | | PR2 <= 0.699
| | | | | PR1 <= 0.696
| | | | | | PR3 <= 0.699
| | | | | | | PR4 <= 0.698: F1 (472.0)
| | | | | | | PR4 > 0.698: S (29.0)
| | | | | | | PR3 > 0.699: S (56.0)
| | | | | | | PR1 > 0.696: S (81.0)
| | | | | | PR2 > 0.699: S (227.0)
| | | | | PR5 > 0.698: S (1540.0)
| | Upv_Soll > 786.5
| | | TIMESTAMP hora <= 38700: E1S1 (93.0)
| | | TIMESTAMP hora > 38700
| | | | Uac <= 237.4: T (132.0/2.0)
| | | | Uac > 237.4
| | | | | pvgis mes <= 2: T (4.0)
| | | | | pvgis mes > 2: E1S1 (11.0)
| PR > 0.699: OK (26226.0)

```

Figura 2. Comparación entre el rendimiento de un generador fotovoltaico de gran escala con el rendimiento de cuatro de los grupos de módulos que lo componen

Los resultados obtenidos muestran que el principal parámetro que decide si el rendimiento es aceptable o no es el rendimiento o *performance ratio*, como era de esperar. Cuando el rendimiento es igual o mayor de 70% puede considerarse que el grupo de módulos funciona correctamente.

Sin embargo cuando el rendimiento es inferior al 70%, desde el punto de vista del experto, algo está obstaculizando el correcto funcionamiento de ese grupo de módulos del generador.

El salto que se encuentra a continuación en el árbol “Upv_Soll > 786.5” ha sido considerado por el experto como un error, seguramente generado por algunos registros de datos erróneos y que pasaron el filtrado previo. El árbol resultante ha interpretado estos registros como un fallo por malas condiciones atmosféricas.

El salto con la condición sobre la irradiancia difusa “Gd <=28” puede ser explicado, puesto que la

irradiancia difusa es baja cuando la directa es alta. Gd es un parámetro estándar que ha sido facilitado por PVGIS [6]. Esta comparación indica que cuando la irradiancia media es alta y el rendimiento es bajo, algo puede estar afectando al rendimiento de los módulos. Cuando el rendimiento de los módulos es bajo también los días anteriores, el algoritmo interpreta que los módulos están sucios o una sombra afecta esa área.

4. Conclusiones

Los algoritmos de aprendizaje computacional han sido ampliamente utilizados para estudiar y predecir la producción de electricidad a partir de la energía solar y para detectar fallos y pérdidas debido a fallos en su funcionamiento.

Un árbol de decisión fue obtenido a partir de la aplicación del algoritmo de aprendizaje C4.5, el cual se entrenó con más de 30,000 registros procedentes de uno de los grupos de módulos que forman un gran generador fotovoltaico de 222 kWp, junto con información procedente de sensores medioambientales.

Referencias

- [1] Y. Wang, P. Zhang, W. Li, N.H. Kan'an, Comparative analysis of the reliability of grid-connected photovoltaic power systems, in: 2012 IEEE Power Energy Soc. Gen. Meet., 2012: pp. 1–8.
- [2] R. Baños, F. Manzano-Agugliaro, F.G. Montoya, C. Gil, A. Alcayde, J. Gómez, Optimization methods applied to renewable and sustainable energy: A review, *Renew. Sustain. Energy Rev.* 15 (2011) 1753–1766.
- [3] M. Chaabene, M.B. Ammar, A. Elhajjaji, Fuzzy approach for optimal energy-management of a domestic photovoltaic panel, *Appl. Energy.* 84 (2007) 992–1001.
- [4] C. Ben Salah, M. Chaabene, M. Ben Ammar, Multi-criteria fuzzy algorithm for energy management of a domestic photovoltaic panel, *Renew. Energy.* 33 (2008) 993–1001.
- [5] I.H. Witten, E. Frank, *Data mining: practical machine learning tools and techniques with Java implementations*, 3er edition, Morgan Kaufmann, San Francisco, Calif., 2011.
- [6] European Commision, *Photovoltaic Geographical Information System (PVGIS)*, (2011).