DATA MINING METHODS FOR FAILURE CLASSIFICATION ON PV-MODULES MONITORED UNDER FIELD-CONDITIONS

G. Behrens[1], A. Dercho[1], H. Quakernack[1], T. Wächter[1], S. Hempelmann[2], I. Kruse[2]
[1]Solar Computing Lab - University of Applied Sciences Bielefeld, Artilleriestraße 9, 32427 Minden, Germany
[2]STORM Energy GmbH, Ludwig-Feuerbach-Str. 69, 90489 Nürnberg, Germany
Telephone +49.571.8385-184, Email: gbehrens@fh-bielefeld.de

ABSTRACT: The focus of this work is to demonstrate the advantage of collecting modul based PV-power system monitoring data with modern communication technology and the intelligent data analysis algorithms of computer science. Prediction and recognition of different faults in large PV-arrays are very important for the effective operation of PV-plants. Different PV-Systems are monitored with a large amount of data collected. The monitoring data is combined with weather data from free internet sources. The challenge in this work is using intelligent data analysis methods of data mining technology for a fault detection of different failure types in real PV-plants. The focus is on using neural network classifier for failure types of partial shading, defect bypass diodes, hailstorm damage and mechanical damage.
Keywords: Monitoring, Performance, PV-Modules, Failure Classification, Data Mining

1 PURPOSE OF THE WORK

Since PV became a significant part of world energy production, more effective monitoring is needed to ensure reliability and a fast and precise method of detecting faults in PV fields. Some photovoltaic systems are equipped with sensors for measuring voltage, power, irradiance and partially module temperature, like it is done by SunSniffer®-Sensor which is used and described by Kilper et al in [1]. Sensor data is normally sent to server collecting all information. This data can be stored and analyzed with the aim to predict module performance [2] and to detect malfunctioning systems automatically. Data Mining Technologies are able to analyze huge amount of time series of monitoring data to recognize and predict faults on the base of large data amounts taken by monitoring PV-plants, what was shown from Braun et al [3]. Marion et al [4] provides a mathematical model of PV-modules performance prediction and tried for the first time to apply Machine Learning algorithms on this field. Based on time series of module based data taken under field conditions by SunSniffer®-Sensor a recognition of failures caused by partial shading effects in PV-fields is possible with high accuracy, as we had shown in our previous work [5]. This work is focused on Data Mining Methods using free WEKA-libraries from University of Waikato [6] with the aim to distinguish between different typical failure types on PV-modules in field. We combined module based monitoring data from SunSniffer®-Sensor with time series of weather data from free web-based interface Forecast.io [http://forecast.io]. Based on the extended information in consolidated long term data series we developed a new data mining application, which is able to classify three different failure types caused by: defect bypass diode, by hailstorm and by simulated mechanical destruction.

2 APPROACH

Typical Data Mining Projects consist of for steps: **Data Collection**, **Data Preprocessing**, **Data Analysis** and **Data Classification**.
In the first step of **Data Collection** measurements are done with sensors and data sets are collected from other available sources. In our application we wrote a web based data interface to data server of SunSniffer®-measurements from the firm StormEnergy. We took data from all modules of three PV-plants over a year. In two of the PV-plants failures on PV-modules like shadowing and defect bypass diodes are known. In one special PV-plant for testing (as shown in figure 1) we could provide some experiments for failure simulation.

Figure 1: Test-site for Failure simulation on PV-modules in Isseroda, Germany. PV- modules are from Scheuten on the top (1S-10S) and PV-moduls from BP Solar (1BP-10BP). All modules are equipped with SunSniffer®-Sensors.

We destroyed on two modules the bypass diodes, damaged one module like a hailstorm, the other like a man went for a walk over the PV-modul by accident. The other PV-moduls worked very well.
The SunSniffer®-sensors (figure 2) measure every 10 minutes for every module the module voltage and the module temperature. Also String current values are collected by SunSniffer®- collection box and sent to data server.

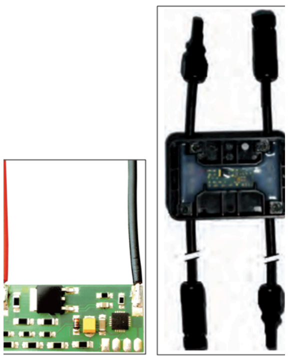

Figure 2: SunSniffer®-Sensor on left side and Sunsniffer®-Retrofit on the right side, as it was mounted on PV modules of experimental test site in Isseroda.

Additionally we imported the values forecast.io like temperature, cloud covering, timestamp of sunrise and sunset and solar efficiency. Value of solar efficiency is calculated as the arithmetic product of the time between sunrise and sunset multiplied with the difference of one minus cloud covering. Forecast.io obtains weather data beginning with the day of 01.01. 1970 and provides up to 1000 answers for free a day. The request URL is shown in figure 3.

```
Forecast.io

HTTP-GET Request
Request-URL:
https://api.forecast.io/forecast/APIKEY/LATITUDE,LONGITUDE,UNIXTIMESTAMP
• APIKEY      = User Verification
• LATITUDE    = degree of latitude
• LONGITUE    = degree of longitude
• UNIXTISTAMP = time in seconds since
    01.01.1970

HTTP Response:
JSON-Object with weather data of one
day in the past.
```

Figure 3: HTTP Request and response for weather data interface from Forecast.io

All Data of long term measurement are stored in Universal Solar database of the Solar Computing Lab as described in [7].

Data collection in step two means also data consolidation from different sources. It was done by database programming an SQL-question for matching the time stamps of measurement series approximately. Generally it is important in Data mining to collect all possible data, which could have some information about the failure events on PV-modules, we want to classify in automatized applications. Data Mining Applications always can give only results as good, as the learning data sets include information about classification aims.

In the second step of **Data Preprocessing** we filtered out data outliers, implemented smoothing algorithms for new calculation of empty and incorrect measuring values, and we normalized all data types into comparable accuracy and value ranges.

For example temperature measurements obtained outliers of -273ºC, which were filtered out and replaced by arithmetic middle of the next neighbors with “normal” values (figure 4):

$$x_f^{(i)} = \frac{x_{f-1}^{(i)} + x_{f+1}^{(i)}}{2}$$

Figure 4: Interpolation by using next neighbors

Results of interpolation of module temperatur values are shown in figure 5.

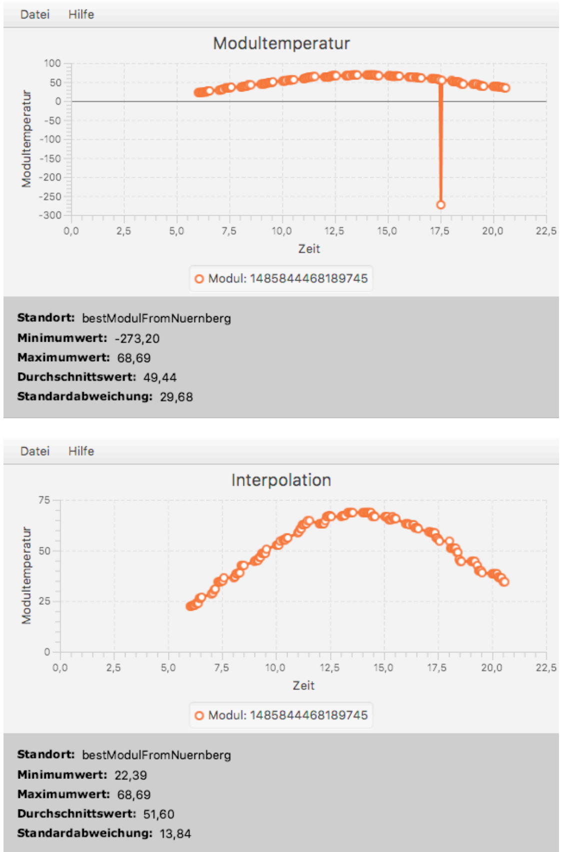

Figure 5: Modul temperature values before (upper chart) and after interpolation operations (lower chart).

Normalization we provided in an easy way by calculation the reciprocal of every value type and received an equal value range of [0.0 , 1.0]. For example modul temperature values after normalization are shown in figure 6.

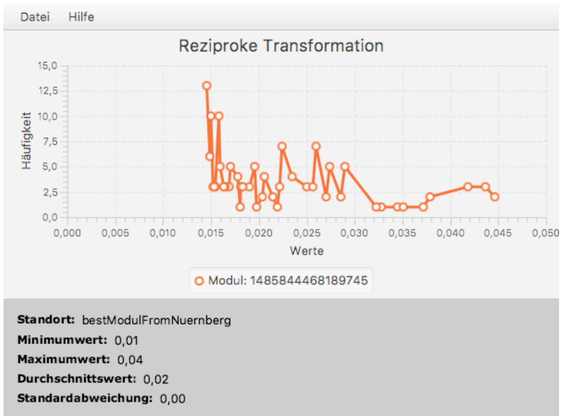

Figure 6: chart of normalized module temperature values

In third step of **Data Analysis** data sets will formed in different feature vector sets, which will be analyzed and optimized.

Feature vectors will be analyzed with focus on statistical correlations between different data types and clusters of data types with statistical correlations between their values with the aim to find out from the measurements the significant data types for failure classification. Values of correlation matrix between all value types are calculated as shown in figure 7:

$$c_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k^{(i)} - \overline{x}^{(i)}\right)\left(x_k^{(j)} - \overline{x}^{(j)}\right), \quad mit \quad i,j = 1,..,p$$

Figure 7: Correlation matrix calculation function, where ***n*** is the number of feature vectors and ***p*** the number of components of feature vector.

For example the correlation between modul temperatur and weather temperature was the highest with an value over the full data sets by 82,02. The twice biggest correlation we calculated between modul temperatur and string current with a value of 43,98. The third biggest correlation was found between modul temperatur and module voltage. A chart showing correlation between modul temperatur and string current visualized for an exemplary PV-modul is given in figure 8.

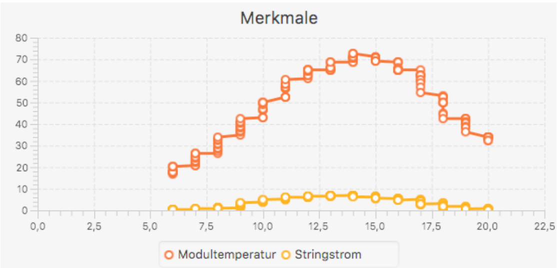

Figure 8: Chart with correlation example of values for module temperature and string currency.

The resulting feature vector of our application after correlation optimization consist of the six following components: modul temperature, modul voltage, string current, weather temperature, solar efficiency and cloud covering.

In the fourth step of **Data Classification** the data set was labeled for different failure event types: no failure, defect bypass diode, hailstorm, simulated mechanical destruction. Different machine learning algorithms from WEKA-Library [6] are applied and parameterized in a huge amount of calculating experiments on the manually labeled data sets. Best results we got on neural network classifier, which we parameterized experimentally in the parameters of the epochs, the momentum, the learning rate and the ratio of the number of feature vectors in training data dividing by the number of feature vectors of the testing data sets.

We investigated mainly three classification experiments. The first one trained into four failure types (top reference module, defect bypass diode, hailstorm, mechanical destroyed by walking over), the second one was trained for the two classes (top module, all defect module types) and the third experiment trained on the four failure types, but feature vectors only for good sunny conditions (string current > 5,9 A) were taken into training data set. All parameters of neural network classifier were permutated in their value ranges and classification results were written in a data file. Afterwards the data files were analyzed for getting the best parameter set. For example the first experiment the best parameter sets are given in table 1, the corresponding classification accuracies are given in figure 9. The classification results for third experiment are exemplarily given in figure 10. A accuracy of classification by 96% could be achieved.

Table 1: Neural Network parametrization values for best classification results of a classification experiment in all four different failure types

Right classified	Epochs	Momen-tum	Learning Rate	Training Data Ratio
75,21%	572	0,6	0,9	87%
75,15%	695	0,4	1,0	85%

```
=== Confusion Matrix ===

  a   b   c   d    <-- classified as
194  4  14   1 |   a = Destroyed
 70 199 13   0 |   b = Topreference
 44   0  18   1 |   c = Hailstormed
  0   0   0  35 |   d = Defect diode
```

```
=== Confusion Matrix ===

  a    b    c   d    <-- classified as
235   4    2   0 |   a = Destroyed
101 230    0   0 |   b = Topreference
 56   6    9   1 |   c = Hailstormed
  0   0    0  40 |   d = Defect diode
```

Figure 9: Classification results for the neural network classifier trained on four different failure types.

The classification results for third experiment are exemplarily given in table 2 and figure 10. An accuracy of classification by 96% could be achieved, if the conditions for training data sets with 66% are realistic to adopt in an real PV-plant monitoring.

Table 2: Neural Network parametrization values for best classification results of a classification in all four different failure types by taking only sunny conditions under account.

Right classified	Epochs	Momen-tum	Learning Rate	Training Data Ratio
100%	584	0,1	0,1	98%
96,88%	69	0,2	0,5	66%

```
=== Confusion Matrix ===

a b c   <-- classified as
2 0 0 | a = Topreference
0 0 0 | b = Destroyed
0 0 2 | c = Defect diode
```

```
=== Confusion Matrix ===

 a  b   c    <-- classified as
27  1   0  |  a = Topreference
 1 17   0  |  b = Destroyed
 0  0  18  |  c = Defect diode
```

Figure 10: Classification results on four different failure types by taking only sunny conditions into calculation.

3 SCIENTIFIC INNOVATION AND RELEVANCE

The scientific innovation of our work is that our application provides Data Mining Algorithms based on a huge amount of big data combined from modul based monitoring data collected with SunSniffer®-Sensor and webbased source of weather data. The software application is easy to adopt to any monitoring system with the possibility of data collection and failure labeling. The recognition rates of our data mining algorithm process for four different tested failure types are high with 96% accuracy. We could show, that it is possible to recognize different failure types in PV-plants in fully automatic application based on different data types.

4 RESULTS AND CONCLUSIONS

Modul based monitoring of PV-modules by SunSiffer®-Sensor gives the possibility to recognize different failure types like shadowing, defect bypass diodes, hailstorm damage and mechanical damage in PV-Plants. The implementation of Data Mining technology can realize it. The results would be better, than more labeled failure types the algorithm could use for training. On very huge training data sets we can apply deep learning algorithms, which can achieve recognition rates of failure types of nearly 100%.

5 REFERENCES

[1] T. Kilper, I. Kruse, C. Feser, U. Kirstein, D. Peters, M. Tapia, K. Maydell, S. Yilmaz, G. Behrens “A new generation of PV Monitoring system with high-grade remote diagnostics based on module level monitoring and integrated yield simulation” Proceedings of 32th European Photovoltaic Solar Energy Conference and Exhibition, Hamburg 2015

[2] Herbort, Brecht, te Heesen. Insolation dependent solar module performance evaluation from PV monitoring data. Proceedings of 27th European Photovoltaic Solar Energy Conference and Exhibition, Frankfurt 2012, pp.4154-4157

[3] H. Braun, S.T. Buddha, V. Krishnan, A. Spanias, C. Tepedelenlioglu, T. Yeider and T.Takehara. “Signal processing for fault detection in photovoltaic arrays.” In 37th IEEE International Conference on Acoustics, Speech and Signal Processing, pages 1681-1682, 2012

[4] B. Marion. “Comparison of predictive models for PV module performance.” In 33rd IEEE Photovoltaic Specialist Conference, pages 1-6, 2008

[5] G. Behrens, A. Domnik, S. Hempelmann, S. Yilmaz, I. Kruse “Maschine Learning Methods for partial Shading Detection in Monitoring Data on PV-System”, Proceedings of 31th European Photovoltaic Solar Energy Conference and Exhibition, Amsterdam 2014

[6] I.H. Witten, E. Frank, M. A. Hall “Data Mining: Practical Machine Learning Tools and Techniques” Elsevier Inc. 2011

[7] G. Behrens, F. Hamelmann, L. Niemann, J. Weicht, S. Yilmaz, I. Kruse „Universal location based data sets for monitoring and scientific data analysis on PV-systems”, Proceedings of 30th European Photovoltaic Solar Energy Conference and Exhibition, Paris 2013