Lukas PROKOP
Stanislav MISAK
Vaclav SNÁŠEL
Pavel KRÖMER
Jan PLATOŠ

PHOTOVOLTAIC POWER PLANT POWER OUTPUT PREDICTION USING FUZZY RULES

ABSTRACT *Photovoltaic Power Plants (PVPP) are classified as power energy sources with non-stabile supply of electric energy. It is necessary to back up power energy from PVPP for stabile electric network operation. We can set an optimal value of back up power energy with using a variety of prediction models and methods for PVPP Power output prediction. Fuzzy classifiers and fuzzy rules can be informally defined as tools that use fuzzy sets or fuzzy logic for their operations. In this paper, we use genetic programming to evolve a fuzzy classifier in the form of a fuzzy search expression to predict PVPP Power output.*

Keywords: *Photovoltaic Power Plant, Fuzzy Rules, Prediction*

Ing. Lukáš PROKOP, Ph.D., doc. Ing. Stanislav MISAK, Ph.D.
e-mail: lukas.prokop@vsb.cz; stanislav.misak@vsb.cz

Department of Electric Power Engineering, VSB – Technical University of Ostrava

prof. RNDr. Václav SNÁŠEL, CSc.
e-mail: vaclav.snasel@vsb.cz

Ing. Pavel KRÖMER, Ph.D., Ing. Jan PLATOŠ, Ph.D.
e-mail: pavel.kromer@vsb.cz; jan.platos@vsb.cz

Department of Computer Science, VSB – Technical University of Ostrava

PROCEEDINGS OF ELECTROTECHNICAL INSTITUTE, Issue 258, 2012

1. INTRODUCTION

Owing to energy legislation regulating business in the Czech Republic due to which electricity redemption prices are high and disproportional to those in neighbouring countries, the number of requests to connect wind and photovoltaic power plants into the distribution networks has increased immensely. The enormous growth of the installed capacity of wind and photovoltaic power plants in recent years has had an adverse impact on the electricity supply system in the Czech Republic and other EU countries as well.

This situation, after adopting respective measures and an act at the end of 2010, has attenuated and the increase in new plant construction in 2011 has slowed down. The installed capacity of photovoltaic power plants was 1,958.38 MW (approx. 8% of the installed capacity of the electric supply system of the Czech Republic) as of 1 March 2011 (latest published data).

However, the operation of the plants installed by the end of 2010 will continue to have an adverse impact on the operation of the distribution networks in the years to come as well. Besides the negative impacts on voltage quality such as the increased values of harmonic voltage, total harmonic distortion or flicker perception rate that are largely caused by using semi-conductor technology, the photovoltaic power plants have an adverse impact on the electric supply system through power supply instability caused by the variability of weather conditions in installation sites. These are the sources with large variability of supplied power.

The power supplied by a photovoltaic power station is changing very dynamically as a result of the changes in solar radiation intensity. To eliminate such rapid changes in the volume of supplies of power or complete shutdown of these generation units, a regulation system is used by the network operator for securing stable operation of the network. For the regulation it is necessary to use power that is allocated in power plants and that serves just for the regulation purposes. The size of the regulation output needed depends on the output size of operated power plants. As the power supply from unstable renewable sources changes over time, the calculation of the size of needed regulation output is relatively complex because the change of supplied power from photovoltaic power plants takes a matter of just minutes. This calculation is based on planning power supplies from all sources connected to the electricity supply system, thus also from photovoltaic power plants. As the supply from these sources is unstable, monitoring photovoltaic plants operation is important for planning the size of reserves but the key is mainly the possibility to forecast the power generation from these sources for certain time intervals of future operation, for instance for intervals of 12, 24 or 36 hours. Currently predicting for longer intervals has not had greater importance due to the errors that we make in predicting as a result of the large variability of the factors used for the prediction.

Nowadays a number of mathematical methods are used for predicting electricity generation from such unstable sources. These methods are based on employing e.g. meteorological models, time series, neural networks, statistical methods or fuzzy logic.

The models that are commonly used predict solar radiation energy, which is basically the same as for predicting electric power generation. This happens in a couple of steps or by combining several of the above methods. As an example, the most common procedure shown in [7] can be mentioned, where predicting is divided into two consecutive steps.

In the first stage solar energy is normalized using the model for a so-called clear sky with the aim of creating a stable time line. Whereas the standard methods of linear time lines for predicting solar radiation or generated electric power can be employed subsequently.

At this stage, the quality of the clear sky model is quite crucial. This model is used for dividing solar radiation into the direct and diffusion radiation as the ratio of individual solar radiation components changes depending on the amount of clouds. In the case of a completely clear sky the solar radiation contains a high share of direct radiation, and vice versa when the sky is overcast – the share of direct radiation is minimal and diffusion radiation prevails. The share of individual radiation components is connected with the type of the panels that are used in the photovoltaic power plant as each generation of photovoltaic panels can absorb a different solar radiation component. Monocrystalline panels absorb just the direct component of solar radiation, while polycrystalline panels are able to absorb both components, both the direct and the diffusion component.

In the second stage neural networks, genetic algorithms or fuzzy logic are used, with the possibility of large input variability of these models for the direct prediction of solar radiation energy or generated electric power.

The model presented in this article makes use of the possibilities of genetic programming that is applied for finding the fuzzy classifier by means of which the prediction is made.

2. GENETIC PROGRAMMING FOR POWER OUTPUT PREDICTION

Genetic programming is a powerful machine learning technique from the wide family of evolutionary algorithms. In contrast to other evolutionary algorithms, it can be used to evolve complex hierarchical tree-like structures and symbolic expressions.

In this work, we use genetic programming to evolve fuzzy classifiers for a photovoltaic plant power output prediction. In particular, genetic programming is employed to evolve a symbolic fuzzy classifier that is used to estimate photovoltaic plant power output from current values of attributes read from various sensors (e.g. light intensity sensors, wind speed sensors etc.).

The fuzzy classifier can be seen as a special type of decision tree. In contrast to traditional decision trees, it takes inspiration from fuzzy information retrieval. The fuzzy classifier used in this work uses both, operators and evaluation functions, that are commonly utilized in fuzzy information retrieval.

Genetic programming is a supervised machine learning algorithm that can generate the classifiers from a training data set. Such a fuzzy classifier can be subsequently used for efficient and fast prediction of a values of an output variable, for prediction of product quality, for classification of data samples, and generally to assign labels to data. Importantly, the output of the classifier is a real value, i.e. it can be used to estimate course of a real valued function. For convenience, we will call the predictor “a classifier” in the remainder of the text.

Artificial evolution of fuzzy classifiers is a promising approach to data mining because genetic programming has proven very good ability to find symbolic expressions in various application domains. The general process of classifier evolution can be used to evolve classifiers for different data classes and data sets with different properties. The resulting classifiers can be used as standalone data labeling tools or participate in collective decision in an ensemble of data classification methods.

3. GENETIC PROGRAMMING

Genetic programming (GP) is an extension to genetic algorithms, allowing work with hierarchical, often tree-like, chromosomes with an unlimited length [1, 2]. The GP shares the workflow with genetic algorithms (Fig. 1). It iteratively evolves a population of encoded candidate solutions (chromosomes) that are modified by so called genetic operators so that the goodness of the solutions improves.

GP was introduced as a tool to evolve whole computer programs and represented a step towards adaptable computers that could solve problems without being programmed explicitly [1, 3].

In GP the chromosomes take the form of hierarchical variably-sized expressions, point-labeled structure trees. The trees are constructed from nodes of two types, terminals and functions. More formally, a GP chromosome is a symbolic expression created from terminals t from the set of all terminals T and functions f from the set of all functions F satisfying the recursive definition [3]:

1. $\forall\, t \in T$: t is the correct expression;
2. $\forall\, f \in F$: $f\,(e_1, e_2, \ldots, e_n)$ is the correct expression if $f \in F$ and $e_1, \ldots, e_n$ are correct expressions;
3. there are no other correct expressions.

GP chromosomes are evaluated by the recursive execution of instructions corresponding to tree nodes [3]. Terminal nodes are evaluated directly (e.g. by reading an input variable) and functions are evaluated after left-to-right depth-first evaluation of their parameters.

1. Define objective function
2. Encode initial population of possible solutions as fixed-length binary strings and evaluate chromosomes in initial population using objective function
3. Create new population (evolutionary search for better solutions):
   a. Select suitable chromosomes for reproduction (parents)
   b. Apply crossover operator to parents with respect to crossover probability to produce new chromosomes (offspring)
   c. Apply mutation operator to offspring chromosomes with respect to mutation probability. Add newly constituted chromosomes to new population
   d. Until the size of new population is smaller than size of current population go back to a.
   e. Replace current population by new population
4. Evaluate current population using objective function
5. Check termination criteria; if not satisfied go back to III.

Fig. 1. Genetic algorithms workflow

Genetic operators are applied to the nodes in the tree-shaped chromosomes. A crossover operator is implemented as the mutual exchange of randomly selected sub-trees of the parent chromosomes.

Mutation has to modify the chromosomes by pseudo-random arbitrary changes in order to prevent premature convergence and broaden the coverage of the fitness landscape. Mutation could be implemented as:

1. removal of a sub-tree at a randomly chosen node;
2. replacement of a randomly chosen node by a newly generated sub-tree;
3. replacement of node instruction by a compatible node instruction (i.e. a terminal can be replaced by another terminal, a function can be replaced by another function of the same arity);
4. a combination of the above.

4. FUZZY CLASSIFIER EVOLUTION BY GENETIC PROGRAMMING

We use an algorithm for fuzzy classifier evolution inspired by the principles of fuzzy information retrieval and evolutionary optimization of search queries [4].

The fuzzy classifier takes form of a symbolic expression with data features (data set attributes) as terminals and operators as non-terminal nodes. Both terminals and non--terminals are weighted. An example of the fuzzy classifier is shown in Figure 2.

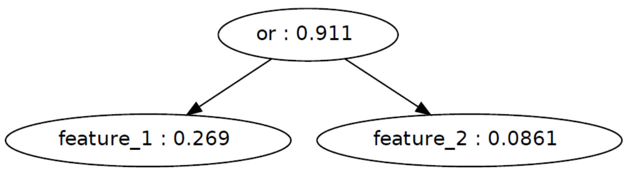

Fig. 2. An example of a fuzzy classifier
feature 1:0.269 or: 0.911 feature 2:0.0861

Fuzzy classifier is evaluated for each data sample in the training collection. For each terminal, the value of corresponding feature is taken. The operators are implemented with the help of standard fuzzy set operators, i.e. x and y is implemented as $min(x,y)$, x or y is implemented as $max(x,y)$, and not x is implemented as $1 - x$. The standard implementation of fuzzy set operators was used but any other pair of t-norm and t-conorm or Ordered Weighted Averaging (OWA) operators could be used.

Classifier weights are used to smoothen the influence of classifier operators and to blur the meaning the data features. Its use allows forming rich and flexible classification statements. There are many ways to interpret and subsequently compute the classifier weights. In this work, the classifier weights are interpreted as threshold (e.g. data samples with feature values greater than the corresponding classifier weight are awarded by greater value) [5]:

$$g(F(d,t),a) = \begin{cases} e^{K(F(d,t)-a)^2} & F(d,t) < a \\ P(a) + Q(a)\dfrac{F(d,t)-a}{1-a} & F(d,t) \geq a \end{cases} \qquad (1)$$

where $P(a)$ and $Q(a)$ are coefficients used for tuning the threshold curve. $P(a)$ and $Q(a)$ used in our implementation are $P(a) = (1+a)/2$ and $Q(a) = (1-a^2)/4$. The other symbols in (1) are: t represents a feature in the data set, d is a data sample, $F(d,t)$ is the value of feature t in data sample d, a is the weight of feature t in the classifier.

The evaluation of a classifier over the training data set assigns to each data record real value from the interval [0;1] which can be interpreted as membership degree of the data record in a fuzzy set defined by the classifier.

The fitness value of the classifier is then evaluated using the information retrieval measure F-score F:

$$F = \frac{(1+\beta^2)PR}{\beta^2 P + R} \tag{2}$$

which is a scalar combination of precision P and recall R. Precision and recall are for two fuzzy sets (pattern A and classifier C) computed using Σ-count:

$$\rho(X,Y) = \begin{cases} \dfrac{\|X \cap Y\|}{\|Y\|} & \|Y\| \neq 0 \\ 1 & \|Y\| = 0 \end{cases} \tag{3}$$

$$P = \rho(A,C) \qquad R = \rho(C,A) \tag{4}$$

5. EXPERIMENTS

The genetic programming was used to evolve a fuzzy rule to estimate the power output of a photovoltaic power plant. We have implemented the genetic programming for fuzzy rule evolution according to the principles outlined above and used it to evolve a classifier that estimates the output power based on the sensor readings. In particular, the genetic programming generated a random set of candidate classifiers. The classifiers had random structure and random weights assigned to the nodes. In the course of the evolution, both the classifier structure and weight values were modified using the genetic operators.

A data set from a real photovoltaic power plant was used to evolve the classifier. The data set contained 24030 records containing values from 2 light intensity sensors and 1 wind sensor. Each record also contained the power output of the plant at given moment. All rows in the data matrix were normalized into the interval [0,1].

Even though the data set contains only three features, the genetic programming is a good way to seek for the dependencies between input data and output value. The algorithm can generate a classifier with complex structure that might contain the same features many times, perhaps with different weights. Moreover, it can discover the way the features need to be combined in order to get good estimate of the output value.

The data set was divided into a training and testing collection. The training and testing collections contained 14416 and 9612 records respectively.

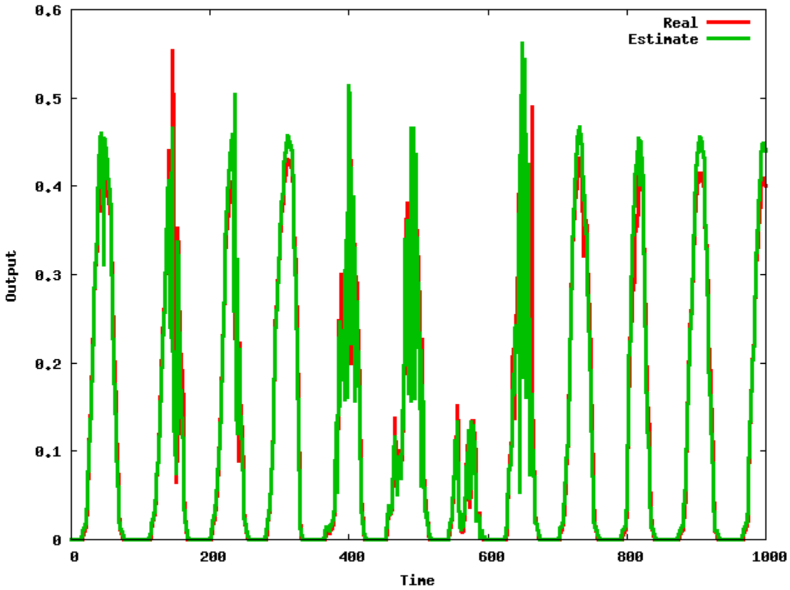

Fig. 3. Real and estimated power output for first 1000 training records

The training collection was used for classifier evolution and the testing collection was used for the evaluation of the evolved classifier. An example of the real and estimated power output for first 1000 training records is shown in Figure 3 and an example of real and estimated power for first 1000 testing records is shown in Figure 4.

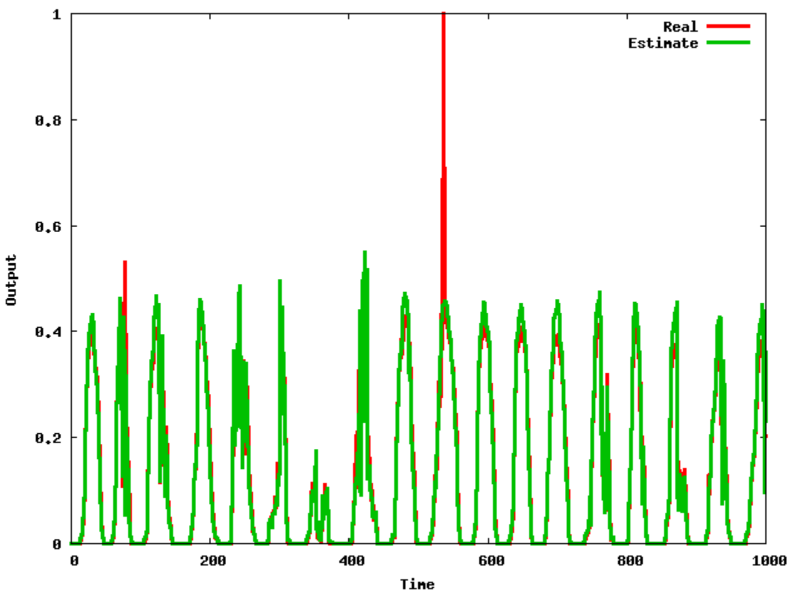

Fig. 4. Real and estimated power output for first 1000 testing records

We can see that the estimated power output corresponds with real power output quite well. The average estimation error for training collection was 0.011 and the average estimation error for testing collection was 0.007. We also note that the data contained some noise (see e.g. the outlying real power output value in Figure 4). The noise affects both, the training and evaluation.

The best classifier found by the algorithm was $Feature_0$:*0.121322*. and several slightly worse classifiers contained only one sensor value as well. It means that the algorithm has repeatedly chosen just one of the light sensors as the most influential input for power output estimation.

6. CONCLUSIONS

Predicting electric power generation is a highly hot topic considering the situation that arose in the Czech Republic as a result of the inappropriately chosen redemption prices of electric power generated in photovoltaic power plants. This legislation has led to large-scale construction of photovoltaic power plants. By the beginning of 2009 only 65.74 MWp were installed, throughout 2009 another approx. 400 MWp were installed, and in the course of 2010 the installed capacity of photovoltaic power plants reached the output of 1952 MWp. Such a volume of installed capacity within the electric supply system of the Czech Republic can cause problems in managing the system in situations when relatively rapid changes in power supply from these unstable sources occur.

One of the possible solutions to this situation is the development and optimisation of the prediction model, which will be able to predict, for the defined time interval of 12, 24 or 36 hours, the volume of electric power that will probably be generated from photovoltaic power plants and thus will enable the distribution system operators to allocate a sufficient amount of regulation power in standard power plants that can participate in regulating the electric supply system.

The precise determination of allocated output has not only technical importance but also economic importance because possible reduction of allocated output size leads to reducing regulation costs.

This paper presents a soft computing method for search for an efficient fuzzy classifier to predict power output of a photovoltaic power plant. The algorithm uses genetic programming and builds on the principles of fuzzy information retrieval. An experimental evaluation has shown that the classifiers found by the algorithm provide reasonable estimate of the photovoltaic plant output power. The results obtained by fuzzy classifier evolution are encouraging. The generic algorithm can be tuned for this application domain and in the future, more soft computing methods for power output estimation can be investigated.

Acknowledgment

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (ENET No. CZ.1.05/2.1.00/03.0069) and project SP2012/53 and also project Czech grant agency No. 102/09/1842.

LITERATURE

1. Koza J.: Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems, Technical Report STAN-CS-90-1314, Dept. of Computer Science, Stanford University, 1990.
2. Koza J.: Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge, MA, USA: MIT Press, 1992.
3. Affenzeller M., Winkler S., Wagner S., Beham A.: Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications. Chapman & Hall/CRC, 2009.
4. Snásel V., Krömer P., Platos J., Abraham A.: The evolution of fuzzy classifier for data mining with applications, in SEAL (K. Deb, A. Bhattacharya, N. Chakraborti, P. Chakroborty, S. Das, J. Dutta, S.K. Gupta, A. Jain, V. Aggarwal, J. Branke, S.J. Louis, K.C. Tan, eds.), vol. 6457 of Lecture Notes in Computer Science, pp. 349-358, Springer, 2010.
5. Prokop L., Misak S., Sikora T., Dasal K., Popławski T., Rusek B.: Optimizing Electric Network Running with Wind Power Plant by using Production Prediction Mathematical Models, Rynek Energii, 2009, vol. II, no. 3, p. 313-318.
6. Misak S., Prokop L.: Prediction System for Energy Production from WPP, In proceedings of 11th International Scientific Conference Electric Power Engineering 2010 (IEEE EPE 2010), 2010, Brno: VUT Brno, pp. 301-305.
7. Bacher P., Madsen M., Nielsen H.A.: Online short-term solar energy forecasting, Solar Energy, vol. 83, pp. 1772-1783, 2009.

Manuscript submitted 18.04.2012

PRZEWIDYWANA MOC WYJŚCIOWA ELEKTROWNI FOTOWOLTAICZNEJ OKREŚLONA PRZY UŻYCIU ZASAD ROZMYTYCH

Lukas PROKOP, Stanislav MISAK, Vaclav SNÁŠEL
Pavel KRÖMER, Jan PLATOŠ

STRESZCZENIE *Elektrownie fotowoltaiczne (EF) są klasyfikowane jako źródła prądu elektrycznego o niestabilnej dostawie energii elektrycznej. Dla stabilnej pracy sieci elektrycznej konieczne jest wspieranie dostawy prądu z EF. Możemy ustalić optymalną wartość* *wspierającej dostawy prądu, stosując różne modele przewidywania i metody dla predykcji mocy wyjściowej z EF. Możliwe jest nieformalne określenie rozmytych klasyfikatorów i zasad jako narzędzi do ich działania, opartych na zbiorach rozmytych i logice rozmytej. W tej pracy stosujemy genetyczne programowanie do opracowania klasyfikatora rozmytego wyrażenia poszukiwania mocy wyjściowej EF.*

Słowa kluczowe: *elektrownia fotowoltaiczna, zasady rozmyte, przewidywanie*