Electrical Power and Energy Systems 60 (2014) 126–140

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods

Pınar Tüfekci *

Department of Computer Engineering, Faculty of Çorlu Engineering, Namık Kemal University, TR-59860 Çorlu, Tekirdağ, Turkey

ARTICLE INFO

Article history:
Received 7 April 2013
Received in revised form 15 February 2014
Accepted 25 February 2014
Available online 24 March 2014

Keywords:
Prediction of electrical power output
Combined cycle power plants
Machine learning methods

ABSTRACT

Predicting full load electrical power output of a base load power plant is important in order to maximize the profit from the available megawatt hours. This paper examines and compares some machine learning regression methods to develop a predictive model, which can predict hourly full load electrical power output of a combined cycle power plant. The base load operation of a power plant is influenced by four main parameters, which are used as input variables in the dataset, such as ambient temperature, atmospheric pressure, relative humidity, and exhaust steam pressure. These parameters affect electrical power output, which is considered as the target variable. The dataset, which consists of these input and target variables, was collected over a six-year period. First, based on these variables the best subset of the dataset is explored among all feature subsets in the experiments. Then, the most successful machine learning regression method is sought for predicting full load electrical power output. Thus, the best performance of the best subset, which contains a complete set of input variables, has been observed using the most successful method, which is Bagging algorithm with REPTree, with a mean absolute error of 2.818 and a Root Mean-Squared Error of 3.787.

1. Introduction

In order for accurate system analysis with thermodynamical approaches, a high number of assumptions is necessary such that these assumptions account for the unpredictability in the solution. Without these assumptions, a thermodynamical analysis of a real application compels thousands of nonlinear equations, whose solution is either almost impossible or takes too much computational time and effort. To eliminate this barrier, the machine learning approaches are used mostly as alternative instead of thermodynamical approaches, in particular, to analyze the systems for arbitrary input and output patterns [1].

Predicting a real value, which is known as regression, is the most common problem researched in machine learning. For this reason, machine learning algorithms are used to control response of a system for predicting a numeric or real-valued target feature. Many real-life problems can be solved as regression problems, and evaluated using machine learning approaches to develop predictive models [2].

This paper deals with several machine learning regression methods for a prediction analysis of a thermodynamic system, which is a combined cycle power plant (CCPP) with two gas turbines, one steam turbine, and two heat recovery systems. Predicting electrical power output of a power plant has been considered a critical real-life problem to construct a model using machine learning techniques. To predict full load electrical power output of a base load power plant correctly is important for the efficiency and economic operation of a power plant. It is useful in order to maximize the income from the available megawatt hours (MW h). The reliability, and sustainability of a gas turbine depend highly on prediction of its power generation, particularly when it is subject to constraints of high profitability and contractual liabilities.

Gas turbine power output primarily depends on the ambient parameters which are ambient temperature, atmospheric pressure, and relative humidity. Steam turbine power output has a direct relationship with vacuum at exhaust. In the literature, the effects of ambient conditions are studied with intelligent system tools such as Artificial Neural Networks (ANNs) for prediction of electrical power (P_E) [1,3,4]. In [1], the effects of ambient-pressure and temperature, relative humidity, wind-velocity and direction on the plant power are investigated using the ANN model, which is based on the measured data from the plant. In [4], the ANN model is used to predict the operational and performance parameters of a gas turbine for varying local ambient conditions.

* Tel.: +90 282 250 23 00; fax: +90 282 652 93 72.
E-mail address: ptufekci@nku.edu.tr

http://dx.doi.org/10.1016/j.ijepes.2014.02.027
0142-0615/

The intelligent systems are also used for modeling a stationary gas turbine. For instance, ANNs identification techniques are developed in [5] and the results show that ANN system identification are perfectly applicable to estimate gas turbine behaviors in wide range of operating points from full speed no load to full load conditions. In [6], Multi Layer Perceptron (MLP) and Radial Basis Function (RBF) Networks are used for identification of stationary gas turbine in startup stage. In [7], dynamic linear models and Feed Forward Neural Networks are compared for gas turbine identification and the neural network is found as a predictor model to identify with better performances than the dynamic linear models. In [8,9], the ANNs models are also used for performance analysis, anomaly detection, fault detection and isolation of gas turbine engines.

Furthermore, in the literature, several studies, e.g., [10–16] have been undertaken to predict electricity energy consumption using machine learning tools, only a few studies such as [1], which is related to prediction of the total electric power of a cogeneration power plant with three gas turbines, one steam turbine and a district a heating system, are found to be as a similar study of this paper.

Pertaining to power plants, it is needed to develop a reliable predictive model for the following day's net energy yield (full load electrical power output) per hour by using real-valued target feature. For this task, there are two main purposes of this study. First one is to determine the best subset of the dataset, which gives the highest predictive accuracy with a combination of the input parameters defined for gas and steam turbines such as ambient temperature, vacuum, atmospheric pressure, and relative humidity. For this purpose, the effects of different combinations of the parameters were investigated and analyzed on predicting full load electrical power output by using 15 machine learning regression methods in WEKA [17] toolbox. Afterwards, the results of the predictive accuracies for the different combinations of the parameters were compared and evaluated to find out the best subset of the dataset. This paper compared the predictive accuracies of the regression methods as the second purpose, which was found out the most successful regression method in the prediction of full load electrical power output of a base load operated CCPP with the highest prediction accuracy.

The remainder of this paper is organized as follows. In Section 2 materials and methods are elaborated, whereas the experimental work is given in Section 3. In Section 4 we provide a discussion of the study and then we conclude in Section 5.

2. Materials and methods

2.1. System description

A combined cycle power plant is composed of gas turbines (GT), steam turbines (ST) and heat recovery steam generators (HRSG). In a CCPP, the electricity is generated by gas and steam turbines, which are combined in one cycle, and is transferred from one turbine to another [18]. A gas turbine in a combined cycle system does not only generate the electrical power but also generates fairly hot exhaust. Routing these gases through a water-cooled heat exchanger produces steam, which can be turned into electric power with a coupled steam turbine and generator. Hence, a gas turbine generator generates electricity and waste heat of the exhaust gases is used to produce steam to generate additional electricity via a steam turbine. This type of power plant is being installed in increasing numbers around the world where there is access to substantial quantities of natural gas [19].

The CCPP,[1] which supplied the dataset for this study, is designed with a nominal generating capacity of 480 MW, made up of 2×160 MW ABB 13E2 Gas Turbines, $2 \times$ dual pressure Heat Recovery Steam Generators (HRSG) and 1×160 MW ABB Steam Turbine as illustrated in Fig. 1.

Gas turbine load is sensitive to the ambient conditions; mainly ambient temperature (AT), atmospheric pressure (AP), and relative humidity (RH). However, steam turbine load is sensitive to the exhaust steam pressure (or vacuum, V). These parameters of both gas and steam turbines, which are related with ambient conditions and exhaust steam pressure, are used as input variables in the dataset of this study. The electrical power generating by both gas and steam turbines is used as a target variable in the dataset. All the input variables and target variable, which are defined as below, correspond to average hourly data received from the measurement points by the sensors denoted in Fig. 1.

(1) *Ambient Temperature (AT):* This input variable is measured in whole degrees in Celsius as it varies between 1.81 °C and 37.11 °C.
(2) *Atmospheric Pressure (AP):* This input variable is measured in units of minibars with the range of 992.89–1033.30 mbar.
(3) *Relative Humidity (RH):* This variable is measured as a percentage from 25.56% to 100.16%.
(4) *Vacuum (Exhaust Steam Pressure, V):* This variable is measured in cm Hg with the range of 25.36–81.56 cm Hg.
(5) *Full Load Electrical Power Output (P_E):* P_E is used as a target variable in the dataset. It is measured in mega watt with the range of 420.26–495.76 MW.

2.2. Feature subset selection

Data pre-processing is a significant process that contains the processes of cleaning, integration, transformation, and reduction of data for using quality data in machine learning (ML) algorithms. Data sets may vary in dimension from two to thousands of features, and many of these features may be irrelevant or redundant. Feature subset selection decreases the data set dimension by removing irrelevant and redundant features from an original feature set. The objective of feature subset selection is to procure a minimum set of original features. Using the decreased set of original features enables ML algorithms to operate faster and more effectively. Therefore, it helps to predict more correctly by increasing learning accuracy of ML algorithms and improving result comprehensibility [20].

The feature selection process begins by inputting an original feature set, which includes n number of features or input variables. At the first stage of feature selection, which is called subset generation, a search strategy is used for producing possible feature subsets of the original feature set for evaluation. Abstractly, the current best subset of the original feature set can be performed by evaluating all the possible feature subsets, which are all the contending 2^n possible subsets. This search is known as exhaustive search, which is too costly and impracticable if the original feature set consists of enormous features [21]. There are also some several search procedures to find the optimal subset of the original feature set, which are more realistic, easier and more practical. However, in this study, the exhaustive search is used as search procedure. Therefore, every feature combination is tried and marked with a score by using ML regression methods, which equals a value of the prediction accuracy. Then the results of each ML regression method are compared to find the feature subset with the best prediction accuracy, which is called as the best subset.

2.3. Machine learning regression methods

A machine learning (ML) algorithm estimates an unknown dependency between the inputs, which are independent variables,

[1] The name of donor power plant is kept confidential.