2014

Optimizing Boiler Efficiency by Data Mining Teciques: A Case Study

Hieu Duong Ngoc
Ho Chi Minh City University of Technology, Vietnam, dnhieu@cse.hcmut.edu.vn

Loc Hoang Xuan
Ho Chi Minh City University of Technology, Vietnam, xuanlochoang2000@gmail.com

Tam Nguyen Minh
Petro Vietnam Fertilizer and Chemical Corporation, Petro Vietnam Group., nmtam@pvfcco.com.vn

Dat Nguyen Huu Quoc
Petro Vietnam Fertilizer and Chemical Corporation, Petro Vietnam Group., nhqdat@pvfcco.com.vn

Hieu Nguyen Quang
Petro Vietnam Fertilizer and Chemical Corporation, Petro Vietnam Group., pqhieu@pvfcco.com.vn

See next page for additional authors

Authors
Hieu Duong Ngoc, Loc Hoang Xuan, Tam Nguyen Minh, Dat Nguyen Huu Quoc, Hieu Nguyen Quang, Hien T. Nguyen, and Vaclav Snasel

22P. Optimizing Boiler Efficiency by Data Mining Teciques: A Case Study

Hieu Duong Ngoc, Loc Hoang Xuan
Ho Chi Minh City University of Technology, Vietnam
dnhieu@cse.hcmut.edu.vn, xuanlochoang2000@gmail.com

Tam Nguyen Minh, Dat Nguyen Huu Quoc, Hieu Nguyen Quang
Petro Vietnam Fertilizer and Chemical Corporation, Petro Vietnam Group.
{nmtam, nhqdat, pqhieu}@pvfcco.com.vn

Hien T. Nguyen§
Ton Duc Thang University, Vietnam
hien@tdt.edu.vn

Vaclav Snasel
VSB-Technical University of Ostrava, Czech Republic
vaclav.snasel@vsb.cz

Abstract

In a fertilizer plant, the steam boiler is the most important component. In order to keep the plant operating in the effective mode, the boiler efficiency must be observed continuously by several operators. When the trend of the boiler efficiency is going down, they may adjust the controlling parameters of the boiler to increase its efficiency. Since manual operation usually leads to unexpectedly mistakes and hurts the efficiency of the system, we build an information system that plays the role of the operators in observing the boiler and adjusting the controlling parameters to stabilize the boiler efficiency. In this paper, we first introduce the architecture of the information system. We then present how to apply K-means and Fuzzy C-means algorithms to derive a knowledge base from the historical operational data of the boiler. Next, recurrent fuzzy neural network is employed to build a boiler simulator for evaluating which tuple of input values is the best optimal and then automatically adjusting controlling inputs of the boiler by the optimal values. In order to prove the effectiveness of our system, we deployed it at Phu My Fertilizer Plant equipped with MARCHI boiler having capacity of 76-84 ton/h. We found that our system have improved the boiler efficiency about 0.28-1.12% in average and brought benefit about 57.000 USD/year to the Phu My Fertilizer Plant.

Keywords

Data Mining, Clustering, K-Means, Fuzzy C-Means, Fuzzy Recurrent Neural Work, Soft Sensor, Boiler

§Corresponding author

1. Introduction

Nowadays, critically fierce competition in modern industrial economy forces companies in industrial sector to seek strategies to reduce cost, increase productivity, and improve production efficiency. An effective strategy will keep those companies growth, efficiency in business, as well as increasing their profits. With a large quantity of good are produced, only increasing by one percent or a few percent of productivity in a year can bring considerable profits. In order to improve production efficiency, one can think about innovative technology, upgrading equipment or optimizing the production process. Observing that the third is the most economical solution and most often applied, we propose an approach applying data mining techniques to optimize production process at Phu My Fertilizer Plant[1], Petro Vietnam Fertilizer and Chemical Corporation, Petro Vietnam Group.

Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years because of the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications, ranging from market analysis, fraud detection, customer retention, to production control and science exploration, especially in industrial sector such as production process optimization.

Regarding to production process optimization, the optimal approaches in literature can be divided into three categories [1]:

- Analyzing production and operating models based on thermodynamics and chemistry.
- Applying soft computing methods such as fuzzy logic, evolutionary computing, artificial neural network, etc to seek optimal solutions.
- Combining approaches mentioned-above.

The drawback of approaches belonging to the first category is the lack of applying automatically analysis tools to solve complex mathematical formulas with many parameters sensitive in the environment changed continuously. Particular in production process optimization at Phu My Fertilizer Plant, the construction of computational and comprehensive combustion models precisely is sophisticated due to many complex expressions that changed under ambient conditions [10]. Moreover, such constructed models are lack of characteristics, considered as input parameters, of combustion process changing over time. As the result the optimization process based on those constructed models causes increasing errors after long time applied.

The approaches belong to the second and third categories have been invested and some applications having functions as soft sensors have been developed in literature. A soft sensor, that may be called a virtual sensor, is a computer software collecting multiple values of parameters correlated with each other in a particular technological process. One can mine the correlation between those parameters to derive knowledge and then that knowledge was exploited to optimize industrial processes as well as forecast for a particular problem. Some notable applications including soft sensors can be listed as follows:

---

[1] http://www.dpm.vn/en

- The first one is developed by Yokogawa Corporation and has four soft sensors for advanced process control. One of those sensors, namely RQE, applies artificial neural network in modeling [5].
- The second one is developed by Emerson Corporation, namely DeltaV, and includes virtual sensors employing artificial neural network (ANN) [3], [4], [9].
- The third is developed by Aspen Technology Inc Corporation. It is a software suite AspenONE® Advanced Process Control for Chemicals including many soft sensors such as: Aspen Inferential Qualities (Aspen IQ) applying Partial Least Square- PLS, Fuzzy PLS, ANN, hybrid of PLS and ANN for data analysis and modeling [6], [9].
- The fourth, namely Kn3, is developed by General Electric GE Corporation. It applies ANN and data mining to build a sensor for optimizing engine factory [13].

In general, one can see a soft sensor as a data mining application combined particular industry knowledge to solve forecasting or estimating problems in which data mining techniques, especially neural networks, are relatively frequently used. The applications mentioned-above have been commercialized and deployed for many years; and also proven its efficiency in terms of economic profits and productivity.

This paper introduces an information system that applies data mining techniques and soft computing to improve the efficiency of the boiler at Phu My Fertilizer Plant. In particular, at Phu My Fertilizer Plant, during the operation, the boiler efficiency varies continuously and sometimes it goes down. The efficiency of a boiler is defined as "the percentage of (heat) input energy is used efficiently to generate steam". When the boiler efficiency goes down, it harms productivity of overall systems. The goal of our proposed system is to detect if the boiler efficiency poses a certain signal of going down in real time, seek the most optimal input values, and adjust controlling inputs by optimal values to increase the boiler efficiency. The optimal values are mined from the operational data in the past of the system. This is a challenging task because the system works in real time and the boiler constantly emits an enormous volume of operational data.

The contribution of this paper is two-fold as follows: (i) building a boiler simulator applying data clustering and recurrent fuzzy neural network to seek the optimal values of parameters of the boiler to increase combustion efficiency by editing the controllable parameters for achieving the best boiler efficiency in particular ambient condition; (ii) building an information system that embeds the boiler simulator built improve the efficiency of the boiler at Phu My Fertilizer Plant. This system saved more than 50 thousands dollar (reduced approximately ~1% operation cost of the boiler) after one year deploying at Phu My Fertilizer Plant, which meets the target of cutting operation cost at that factory.

The rest of this paper is organized as follows. Section 2 presents the architecture of our proposed system and some background concepts relevant to the boiler at Phu My Fertilizer Plant. Section 3 presents how we apply data clustering techniques to derive a knowledge base. Section 4 presents the boiler simulator and the algorithms employed to seek the optimal values (of parameters) giving the best boiler efficiency. Experiments and results are presented in Section 5. Finally, we draw conclusion and perspectives for future work.

2. System architecture and background

In this section, we introduce some basic concepts relevant to the boiler at Phu My Fertilizer Plant and the architecture of our proposed system, as well as explaining in turn each module in it.

2.1 Background concepts

Boiler: Boiler [7], [15] is machine which transfers heat of combustion into the water until the water is boiling and converted into steam. MACHI, an auxiliary boiler at Phu My Fertilizer Plant is kind of "Titanium M" (140 tons / hour) and suitable for fuel gas combustion. The boiler is equipped with an automatic control system having high reliability measuring instruments in order to monitor automatic operation of the furnace and the burner.

Boiler efficiency: how to calculate exactly the efficiency of a boiler plays a very important role in optimizing the production process at Phu My Fertilizer Plant. As mentioned above, the efficiency of the boiler is defined as "*the percentage of (heat) input energy is used efficiently to generate steam*". In literature, there are two methods of assessing boiler efficiency: direct and indirect [7], [14]. We adopt them to our system.

Boiler operational data: The controlling and monitoring system of the boiler continuously reads and stores the values of input parameters such as fan speed, pressure, steam temperature, load capacity, etc. The operational data consists of parameters' values controlling and monitoring the boiler. Some mathematical expressions where these parameters are variables are defined to calculate the boiler efficiency. Among these parameters, several ones are controllable, e.g. water temperature, air flow, etc and the others are uncontrollable, e.g. ambient temperature and humidity of the air, etc. The experts at Phu My Fertilizer Plant define if a parameter is controllable or not based on their experience. The boiler simulator in our system aims at finding which combination of input values is the optimal in increasing combustion efficiency and editing the controllable parameters to achieving better boiler efficiency.

2.2 System architecture

Figure 1 shows our system architecture, namely BEO – Boiler Efficiency Optimization. The system includes some main modules. We present these modules as follows:

Preprocessing data: The module has two main functions (i) reading raw operational data from several separate text files or databases, analyzing their structure and storing them in a unit database (SQL Server DBMS) and (ii) Cleaning the database to make sure it has no errors, outliers, and noisy data. Each record in the database consists of input parameters, load of the boiler, and the boiler efficiency.

Data clustering: Since the system works in real time and the historical operational data is enormous, million records in Phu My Fertilizer Plant in particular, reducing the size of operational data is significant in speeding up the system to meet its real time requirement. Data clustering is to group featured records into the same cluster, and derive a knowledge base consisting of the centers of those clusters.

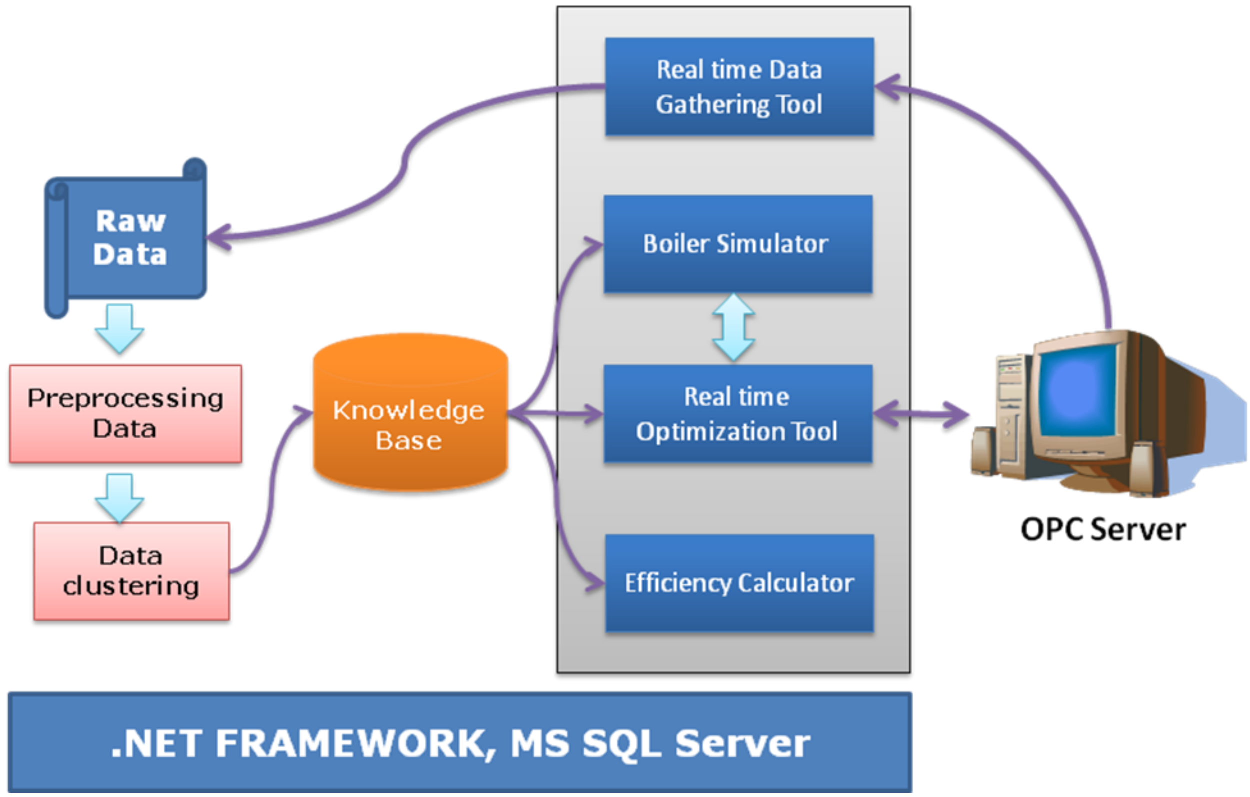

Figure 1: BEO Architecture

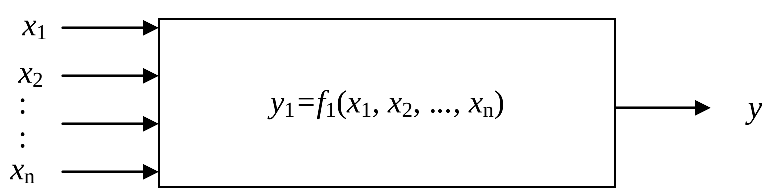

Figure 2: Boiler simulation

Boiler simulator: Combustion is considered as a process of time-oriented technology, and the equation of state combustion efficiency is a complex nonlinear equation where coefficients are not fixed. It is difficult to exactly find coefficients of that nonlinear equation. As a result, we need an approximate solution. Neural networks seem to be a simple and effective approximation scheme for this kind of problem. In our system, the boiler with the internal reaction equation is monitored as a black box, with input parameters and an output. Therefore, we need a boiler simulator for simulating the real boiler by automatically adjusting the tuple of values of the boiler's input parameters in turn and finally choosing the one that gives the best boiler efficiency for the real boiler. The input parameters were collected as: air flow, air pressure, water flow, etc. to supply for the boiler. This boiler is simulated by a multi-variable equation $y = f(x_1, x_2, ..., x_n)$ in which x_i, i=(1..n) are the input parameters as in Figure 2. This modeling of the boiler helps us building a boiler simulation with technological features like a real boiler. The input of the boiler is the current state of all input parameters and the output the boiler efficiency.

After having the boiler simulator, we perform tests to check the accuracy by the boiler simulator. The tests performed on the operational data of the boiler. Then, we will have a variable, namely delta: $\Delta = \mu r - \mu s$, where μr is real efficiency of boiler and μs is the boiler efficiency calculated from the boiler simulator. The simulation results are considered as accuracy if $\Delta \leq \varepsilon$ (ε is an error threshold with a pre-defined value e.g., 0.005 ≈ 0.5%).

Real time optimization tool: At the time of the boiler efficiency appears signal of going down; this module detects in knowledge base and finds the tuple of values of parameters that gives higher efficiency than current efficiency, the controllable parameters will be changed by the new values in the new tuple that has just found. After suggesting the new parameter values, the program predicts the boiler efficiency by boiler simulator module according to these new values. In the case of the predicting efficiency is lower than the current efficiency; the variation will be ignored. Conversely the tool will apply the new found parameter values for the boiler with expecting increasing the boiler efficiency. Process working of the module is illustrated as Figure 4 with several steps as follows:

- *Step 1:* Reading data from OPC Server via Real time data gathering tool.
- *Step 2:* Finding some similar tuples with the current boiler state that has higher efficiency and specially must be the same load.
- *Step 3:* Checking if load changing is higher than a delta value or not. If it is not higher, that means the load is stable and the module can change parameter value to improve boiler efficiency. If it is higher, that means the load is not stable, the module must continue and monitor.
- *Step 4:* In the case of stable load, if the new similar tuple is predicted giving higher efficiency, this will be applied to the boiler in order to improve the boiler efficiency. Otherwise, there is no change on the boiler.

Efficiency calculator: Efficiency can be calculated by the boiler simulator module. Usually boiler simulator only works on *M* important parameters that were chosen by experts. However, data collected at one time is a set of *N* parameters and *N* is usually greater than *M*. There are several formulas for calculating the boiler efficiency based on full *N* parameters for the real boiler. Unfortunately, these are very complex formulas. There are two methods of assessing the boiler efficiency and are used in the BEO system:

- *Direct Method:* As part of the energy obtained from steam compared to the energy of the fuel in the boiler.
- *Indirect Method:* Efficiency is the difference between loss and energy input [7], [14].

Real time data gathering tool: This module is an OPC (OLE for Process Control) Client collecting automatically data from OPC Server via local network. OPC is a software interface standard that allows Windows programs to communicate with industrial hardware devices [17]. In duration time, the tool gets parameters' values from OPC Server and the duration value is defined by the user, usually 60s.

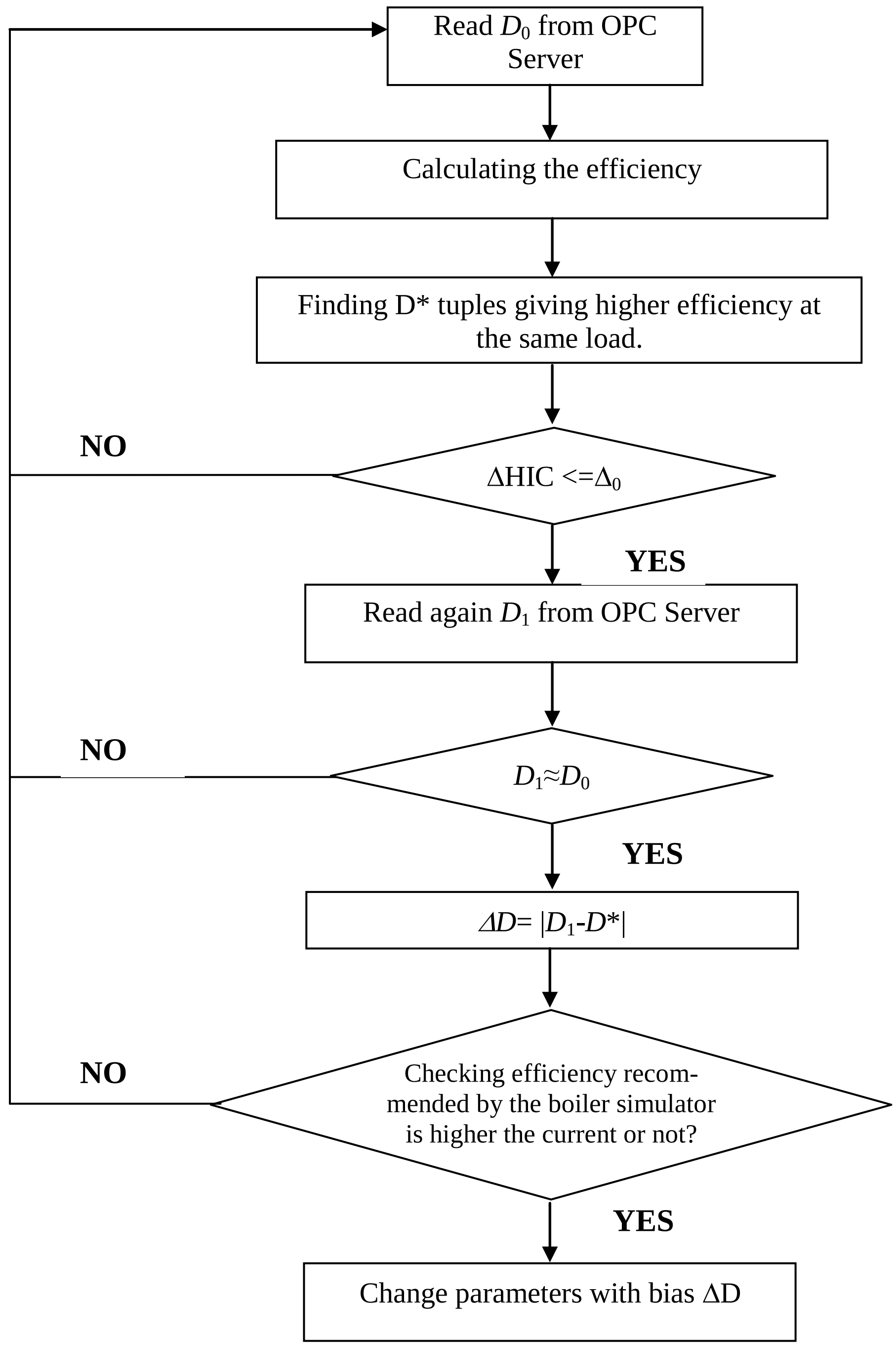

Figure 3: Real time optimization work flow

3. Data clustering

As mentioned above, after pre-processing, we achieve an enormous database containing the historical operational data of the boiler. It consists of records clean up. Each record consists of fields, each of which corresponds to a parameter of the boiler. In particular, in each time slot 60s our system collects a set of data consists of N parameters and the corresponding efficiency and in one year we have 60x24x365 = 525.600 records of data. Since the database is very large, we employ two clustering algorithms: K-means, Fuzzy C-means as two options to reduce its size for speeding up our system' responsibility. The output of data clustering module is a knowledge base that consisting of the centers of result clusters of the clustering algorithms. The knowledge base achieved after running the data clustering module can help our system meet the following targets:

- To speed up the response time: since the boiler works in real time, the response time of the system must be in an acceptable time slot.
- To avoid noisy data: a number of identical rows in a database are grouped and the rows not being grouped are deleted.
- To reduce the size of data but not losing data characteristic.

In the database, we build *n*-dimensional vectors from the records of data. Each dimension of a vector corresponds to a parameter. Each vector is considered as a data point. The center of a cluster C_i shall be calculated according to the following expression:

$c_i = \left(\frac{1}{m_i}\right)\sum P'$ in which m_i is the number of collecting points for cluster C_I and P' are points in cluster C_i.

After clustering, we have many cluster centers that are representative for knowledge base. We call each cluster a signature control. The clustering process has three basic steps:

- Phase 1*:* After preprocessing, vectors will be grouped by different load.
- Phase 2: Using a clustering algorithm, e.g. K-Means or Fuzzy C-Means to create clusters. The value of *K* is determined based on the amount of data point at a particular load level.
- Phase 3: Saving the signature controls calculated to the knowledge base with the condition that these do not exist in the knowledge base.
- Phase 4: Validating the clustering result to decide if stopping or continuing clustering.

4. Boiler simulation

To determine the impact of the set of new optimal parameters that is found by the module real time optimization tool, every time the boiler has signal of efficiency going down, the double checking is very necessary. However, this test will affect the operation of the boiler and requires the approval by a lot of safety requirements and regulations. To this end, we build a boiler simulation to simulate the status of the real system after applying the set of new values of parameters.

There are many methods that can simulate the boiler operation that base on its historical data. We can use regression methods to solve the issue such as linear regression, nonlinear regression, fuzzy system, artificial neural network support vector machine etc. Every method has some advantages and also disadvantages depending on the data with specific characteristics. However, it is not easy to state which method is better than other and the appropriate answer for the most suitable chosen method comes from analysis of experiment results. After evaluating the boiler historical data we decided to choose artificial neural network.

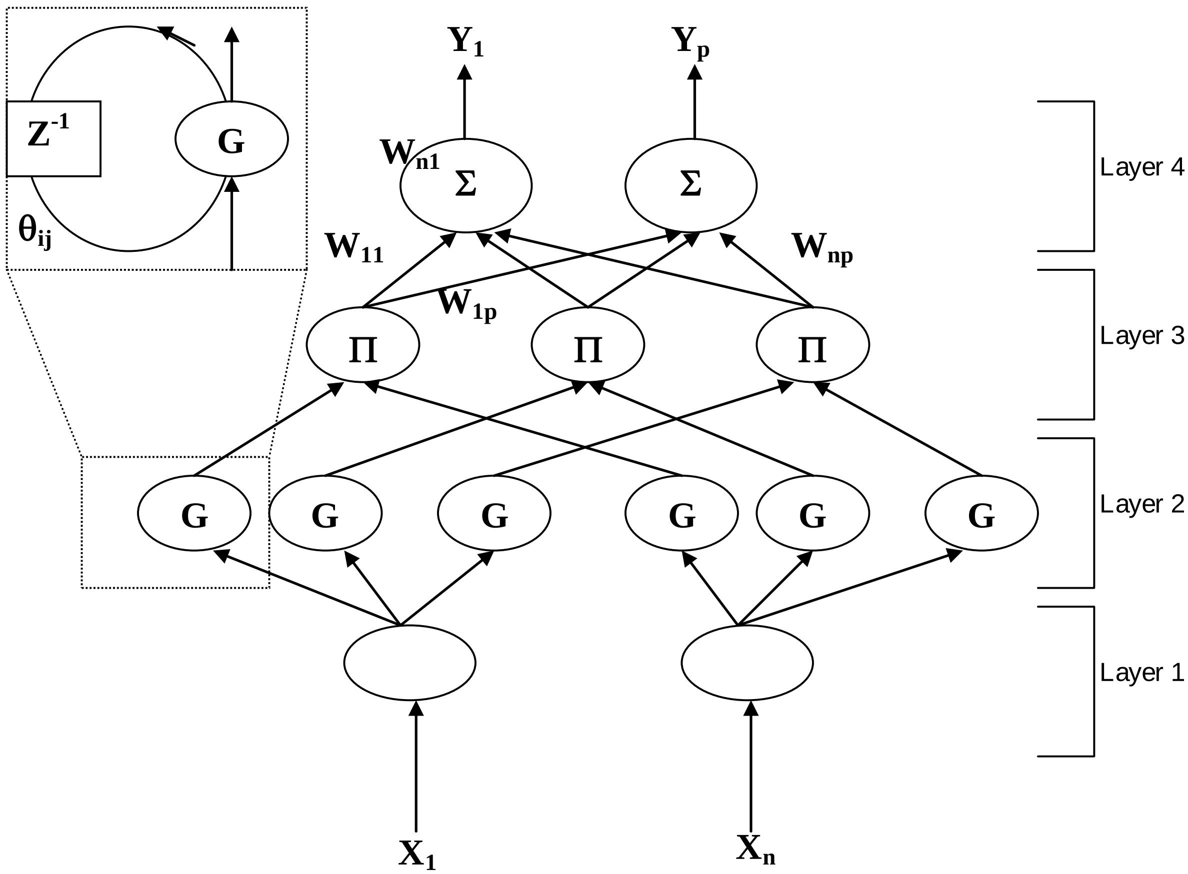

Figure 5: *R*FNN Architecture

As pointed out in [2], the strength of the neural network is that it can approximate any function if the parameters are determined appropriately. If there are available input and output sources, neural networks can model the hidden rule of an unknown system, including the system is linear or non-linear but in strict condition that we have enough data (guaranteeing quantity and quality criterion) [12]. In our system, we employ recurrent fuzzy neural network (RFNN) for simulating boiler operation. This model is hybrid of fuzzy theory and neural network. It works more effective than standard neural network [2], [8]. Structure of RFNN is shown as Figure 5 and RFFF includes four layers as follows:

- Layer 1 is input layer that has *N* nodes, each of which corresponds with a parameter. In our data, input could be all parameters of boiler and depend on the way that we highlight data.
- Layer 2 is called membership layer. Nodes in this layers will convert the crisp data in fuzzy data by applying membership functions such as Gauss function. Number of neural nodes in this layer is *NxM* where *M* is the number of fuzzy rules.
- Layer 3 is the layer of fuzzy rules. Each node in this layer plays the role of a fuzzy rule. Connecting between Layer 3 and Layer 4 presents for fuzzy conclusion.

- Layer 4 is the output layer including *P* nodes. In our model, *P* will be set to one and this is efficiency of boiler.

4.1 *M*odel process

We propose a model that learns parameters m_{ij}, σ_{ij}, θ_{ij} and w_{jk} respectively. Let $u_i^{(k)}$ and $O_i^{(k)}$ be respectively the input and the output of the node i^{th} in the layer *k*.

- Layer 1: $O_i^{(1)} = u_i^{(1)} = x_i(t)$ $\quad$ $i = 1 \div N$
- Layer 2: note that every node has 3 parameters, namely m_{ij}, σ_{ij} and θ_{ij} respectively.

$$O_{ij}^{(2)} = \exp\left[-\frac{\left(u_{ij}^{(2)} - m_{ij}\right)^2}{\left(\sigma_{ij}\right)}\right]$$

(1) $i = 1 \div N$, $j = 1 \div M$, m_{ij} and σ_{ij} are the center and the variance of Gauss distribution function.

$$u_{ij}^{(2)}(t) = O_i^{(1)} + \theta_{ij} O_{ij}^{(2)}(t-1)$$

(2) $i = 1 \div N$, $j = 1 \div M$, θ_{ij} denotes the weight of a recurrent node.

We easily realize that the input of the nodes in this layer has the factor $O_{ij}^{(2)}(t-1)$. This factor denotes the remaining information of the previous model. Therefore, after replacing $u_{ij}^{(2)}$ in (1) by (2), we Equation (3) as follows:

$$\begin{aligned} O_{ij}^{(2)} &= \exp\left[-\frac{\left[O_i^{(1)} + \theta_{ij} O_{ij}^{(2)}(t-1) - m_{ij}\right]^2}{\left(\sigma_{ij}\right)^2}\right] \\ &= \exp\left[-\frac{\left[x_i(t) + \theta_{ij} O_{ij}^{(2)}(t-1) - m_{ij}\right]^2}{\left(\sigma_{ij}\right)^2}\right] \quad (3) \end{aligned} \qquad i = 1 \div N, j = 1 \div M$$

- Layer 3: Each node in this layer corresponds with an AND expression. Each AND expression is defined as follows:

$$\begin{aligned} O_j^{(3)} &= \prod_{i=1}^{N} O_{ij}^{(2)} \\ &= \prod_{i=1}^{N} \exp\left[-\frac{\left[x_i(t) + \theta_{ij} O_{ij}^{(2)}(t-1) - m_{ij}\right]^2}{\left(\sigma_{ij}\right)^2}\right] \end{aligned} \qquad where\ j = 1 \div M$$

- Layer 4: Nodes of this layer are responsible for converting fuzzy to crisp.

$$\begin{aligned} y_k &= O_k^{(4)} \\ &= \sum_{j=1}^{M} u_{jk}^{(4)} w_{jk} \\ &= \sum_{j=1}^{M} O_j^{(3)} w_{jk} \\ &= \sum_{j=1}^{M} w_{jk} \prod_{i=1}^{N} \exp\left[-\frac{\left[x_i(t) + \theta_{ij} O_{ij}^{(2)}(t-1) - m_{ij}\right]^2}{\left(\sigma_{ij}\right)^2}\right] \end{aligned} \qquad where\ k = 1 \div P$$

After defining RFNN architecture in Figure 5, RFNN is learnt by back propagation algorithm. Each learning step spends much time about 1-2 days to learn 80% of all data (5 years data). Then RFNN will be tested in 20% data and evaluated. If it gives prediction results correct, the learning step will be stopped; otherwise the learning is continuous.

4.2 Learning algorithm

Back propagation algorithm is applied for learning RFNN that is the same with feed forward neural network (FFNN). The reason why we choose RFNN is that (i) RFNN can estimate the complex relationship between dependent and independent variables and (ii) the convergence of RFNN is faster than feed forward neural network. The target of the applied algorithm is how to minimize the sum square error (SSE):

$$E = \frac{1}{2}\sum_t \left(y^{(d)}(t) - y\ \ (t)\right)^2 = \frac{1}{2}\sum_t \left(y^{(d)}(t) - O^{(4)}(t)\right)^2$$

in which $y^{(d)}(t)$ is the real boiler efficiency and $y\ \ (t) = O^{(4)}(t)$ is the calculated boiler efficiency from RFNN at the t^{th} tuple. In back propagation algorithm, the parameters are updated as follows:

$W(t+1) = W(t) + \Delta W(t) = W(t) + \eta\left(-\frac{\partial E(t)}{\partial W}\right)$, in which W is a parameter vector of the model and η is the learning rate.

Let $e(t) = y^{(d)}(t) - y\ \ (t)$, we have: $\frac{\partial E(t)}{\partial W} = -e(t)\frac{\partial y(t)}{\partial W} = -e(t)\frac{\partial O^{(4)}(t)}{\partial W}$

Therefore, m_{ij}, σ_{ij}, θ_{ij} and w_{jk} will be updated as:

$$w_{jk}(t+1) = w_{jk}(t) - \eta^w \frac{\partial E}{\partial w_{jk}}$$

$$m_{ij}(t+1) = m_{ij}(t) - \eta^m \frac{\partial E}{\partial m_{ij}}$$

$$\sigma_{ij}(t+1) = \sigma_{ij}(t) - \eta^\sigma \frac{\partial E}{\partial \sigma_{ij}}$$

$$\theta_{ij}(t+1) = \theta_{ij}(t) - \eta^\theta \frac{\partial E}{\partial \theta_{ij}}$$

in which,

$$\frac{\partial E}{\partial w_{jk}} = -e(t)O_j^{(3)}$$

$$\begin{aligned}\frac{\partial E}{\partial m_{ij}} &= -e(t)\sum_{j=1}^{M} w_{jk}\frac{\partial O_j^{(3)}}{\partial m_{ij}} \\ &= -e(t)\sum_{j=1}^{M} w_{jk} O_j^{(3)} \frac{2\left[x_i(t) + O_{ij}^{(2)}(t-1)\theta_{ij} - m_{ij}\right]}{\left(\sigma_{ij}\right)^2}\end{aligned}$$

$$\frac{\partial E}{\partial \sigma_{ij}} = -e(t)\sum_{j=1}^{M} w_{jk} \frac{\partial O_j^{(3)}}{\partial \sigma_{ij}}$$

$$= -e(t)\sum_{j=1}^{M} w_{jk} O_j^{(3)} \frac{2\left[x_i(t) + O_{ij}^{(2)}(t-1)\theta_{ij} - m_{ij}\right]^2}{(\sigma_{ij})^3}$$

$$\frac{\partial E}{\partial \theta_{ij}} = -e(t)\sum_{j=1}^{M} w_{jk} \frac{\partial O_j^{(3)}}{\partial \theta_{ij}}$$

$$-e(t)\sum_{j=1}^{M} w_{jk} \frac{-2\left[x_i(t) + O_{ij}^{(2)}(t-1)\theta_{ij} - m_{ij}\right] O_{ij}^{(2)}(t-1)}{(\sigma_{ij})^2}$$

In summary, our learning model is a supervised learning model with t^{th} tuple $\{(X_t, Y_t)\}$. $X_t = (x_1, x_2 \ldots x_n)$ is a parameter vector and Y_t is its corresponding efficiency. The work flow will be executed as follows:

- Step 1: Feed forward X_t through RFFN to get output as:

$$y_k = \sum_{j=1}^{M} w_{jk} \prod_{i=1}^{N} \exp\left[-\frac{\left[x_i(t) + \theta_{ij} O_{ij}^{(2)}(t-1) - m_{ij}\right]^2}{(\sigma_{ij})^2}\right]$$

- Step 2*:* Calculating error between the real value produced by the real system and output of RFNN for s^{th} tuple. Let $e(t)$ be the error value and $e(t) = y_k(t) - y(t)$, in which $y_k(t)$ denotes the output value of RFNN and $y(t)$ denotes the real value.
- Step 3*:* Updating m_{ij}, σ_{ij}, θ_{ij} and w_{jk} as the following expression:

$$w_{jk}(t+1) = w_{jk}(t) - \eta^{w} \frac{\partial E}{\partial w_{jk}}$$

$$m_{ij}(t+1) = m_{ij}(t) - \eta^{m} \frac{\partial E}{\partial m_{ij}}$$

$$\sigma_{ij}(t+1) = \sigma_{ij}(t) - \eta^{\sigma} \frac{\partial E}{\partial \sigma_{ij}}$$

$$\theta_{ij}(t+1) = \theta_{ij}(t) - \eta^{\theta} \frac{\partial E}{\partial \theta_{ij}}$$

5. Experiment

For experiment, we employ a statistic method called statistical inference for two samples [16] to assess how efficiency BEO system brings to the factory. Note that, to calculate confidently, this method requires that the sample size must be large enough so that Student's t-distribution comes close to the normal distribution. We collected operational boiler data with the duration of 1 sample/60s.

Data used for assessment was gotten in total 17 days with two separate periods: (i) 1st period: from 02-11-2013 to 07-11-2013, sample size is 5892 and boiler load distributes from 76 ton/h to 83 ton/h. (ii) 2nd period: from 08-11-2013 to 18-11-2013, sample size is 8834 and boiler load distributes from 72 ton/h to 84 ton/h. We defined minimum sample size is 30 samples at every boiler load. At the time of collecting data, boiler operation was in auto mode and boiler load value is continuous. Therefore, we must round many boiler load values to get one rounded boiler load value. For example: all boiler loads from 68.5 ton/h to 69.49 ton/h are rounded to 69 ton/h. After collecting data, we employ the statistical inference method to prove that applying BEO on boiler operation improve the performance of power consumption.

As mentioned in [16], Student's t-distribution is one of normal distribution families. In Student's t-distribution, mean of n observed data $\bar{x}$ is estimated as below:

$$\bar{x} = \frac{\sum_1^n x_i}{n}$$

And standard deviation σ:

$$\sigma = \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{n-1}}$$

We define some factors that are used in Figure 6 and Figure 8:

- Mean of power consumption when running boiler with BEO applied is $\bar{x}^{BEO}$
- Standard deviation of power consumption when running boiler with BEO applied is $\sigma_{\bar{x}}^{BEO}$
- Mean of power consumption when running boiler without BEO is $\bar{x}^{noBEO}$
- Standard deviation of power consumption when running boiler without BEO applied is $\sigma_{\bar{x}}^{noBEO}$

Power consumption improvement by boiler load

Figure 7 & 8 show the improvement of power consumption between applied BEO and without BEO in the first period. Figure 9 & 10 show the improvement of power consumption between applied BEO and without BEO in the second period. The experiment results show that BEO improved the performance of power consumption with confidence larger 95% at nearly all load. At 80 ton/h boiler load, BEO achieves the improvement of the performance of power consumption with confidence 94.18%. The Figure 7 & 9 show that BEO achieves good improvement in power consumption for boiler load below 78 ton/h and over 82 ton/h. Total improvement of power consumption in observed time of 17 days can be sumarized as follows:

- Power consumption in total while applying BEO = 23167.32 MMBTU
- Power consumption in total while without applying BEO (equivalent calculation) = 23288.63 MMBTU

The experiment results show that the power consumption is reduced in total $\frac{|23167.32-23288.63|}{23288.63} = 0.52\%$. It means that boiler with the support of BEO system improve the performance of power consumption 0.52%.

Load	With BEO		No BEO		Alternative hypothesis H1 confidence			Energy consumption improvement quantity %
	Energy consumption MMBTU/T $\bar{x}^{BEO}$	Standard Deviation $\sigma_{\bar{x}}^{BEO}$	Energy consumption MMBTU/T $\bar{x}^{noBEO}$	Standard Deviation $\sigma_{\bar{x}}^{noBEO}$	Test Statistic Z_o	$Z_{(1-\alpha_{min})}$ satisfies $-Z_{(1-\alpha_{min})} \approx Z_o$ $Z_{(1-\alpha_{min})}$	Alternative hypothesis H1 true with confidence $(1-\alpha_{min}) \times 100\%$	
76	2.73603933	0.008538797	2.83275541	0.00683683	-8.8417	8.8417	100.00	3.41
77	2.73761938	0.004225423	2.77537323	0.00695109	-4.6411	4.6411	100.00	1.36
78	2.72698483	0.002505412	2.74504508	0.00213734	-5.4841	5.4840	100.00	0.66
79	2.73320109	0.001292799	2.73955079	0.00101060	-3.8696	3.8695	99.99	0.23
80	2.7367319	0.001105821	2.74136221	0.00066019	-3.5952	3.5952	99.98	0.17
81	2.73244367	0.001445864	2.74462662	0.00077038	-7.4364	7.4363	100.00	0.44
82	2.73281713	0.001696231	2.74793355	0.00126849	-7.1369	7.1368	100.00	0.55
83	2.73721923	0.002362635	2.74594183	0.00240384	-2.5879	2.5878	99.52	0.32

Figure 6: Data of improvement of power consumption from 02-11-2013 to 07-11-2013

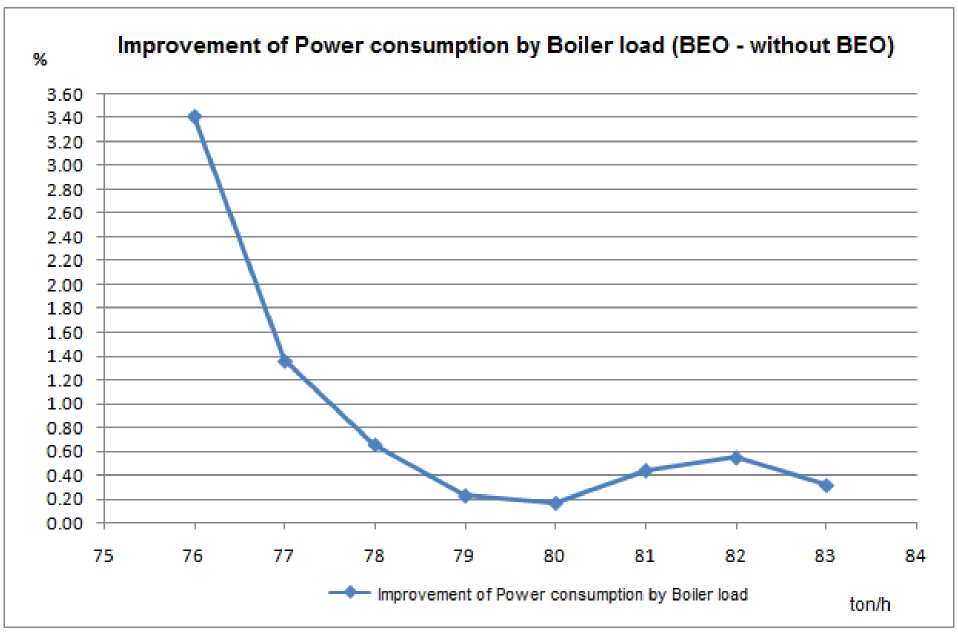

Figure 7: Plot of improvement of Power consumption from 02-11-2013 to 07-11-2013

Load	With BEO		No BEO		Alternative hypothesis H1 confidence			Energy consumption improvement quantity %
	Energy consumption MMBTU/T $\bar{x}^{BEO}$	Standard Deviation $\sigma_{\bar{x}}^{BEO}$	Energy consumption MMBTU/T $\bar{x}^{noBEO}$	Standard Deviation $\sigma_{\bar{x}}^{noBEO}$	Test Statistic Z_o	$Z_{(1-\alpha_{min})}$ satisfies $-Z_{(1-\alpha_{min})} \approx Z_o$ $Z_{(1-\alpha_{min})}$	Alternative hypothesis H1 true with confidence $(1-\alpha_{min}) \times 100\%$	
72	2.71527554	0.003884655	2.768653512	0.00263505	-11.3714304	11.3714000	100.00	1.93
73	2.72383715	0.002091799	2.755741757	0.001817481	-11.5134373	11.5134000	100.00	1.16
74	2.72477586	0.001643908	2.749370391	0.001163027	-12.2134853	12.2134000	100.00	0.89
75	2.72900881	0.001633969	2.750293418	0.000847099	-11.5645926	11.5645000	100.00	0.77
76	2.73108867	0.001807608	2.753405291	0.0007765	-11.3435963	11.3435000	100.00	0.81
77	2.73184048	0.002067067	2.749418841	0.000837597	-7.8815374	7.8815000	100.00	0.64
78	2.73470057	0.001581815	2.753100963	0.000844712	-10.2610281	10.2610000	100.00	0.67
79	2.74096517	0.001106636	2.753691615	0.00120037	-7.7949839	7.7949000	100.00	0.46
80	2.74455284	0.001184484	2.747897977	0.001771387	-1.5698082	1.5698000	94.18	0.12
81	2.73922722	0.001432851	2.749539065	0.002423388	-3.6627960	3.6627000	99.99	0.38
82	2.73677383	0.001302357	2.751834965	0.002105196	-6.0841354	6.0841000	100.00	0.55
83	2.74342237	0.001463261	2.755530878	0.001971886	-4.9311829	4.9311000	100.00	0.44
84	2.74808924	0.00174197	2.767773787	0.003681876	-4.8327410	4.8327000	100.00	0.71

Figure 8: Data of improvement of Power consumption from 08-11-2013 to 18-11-2013

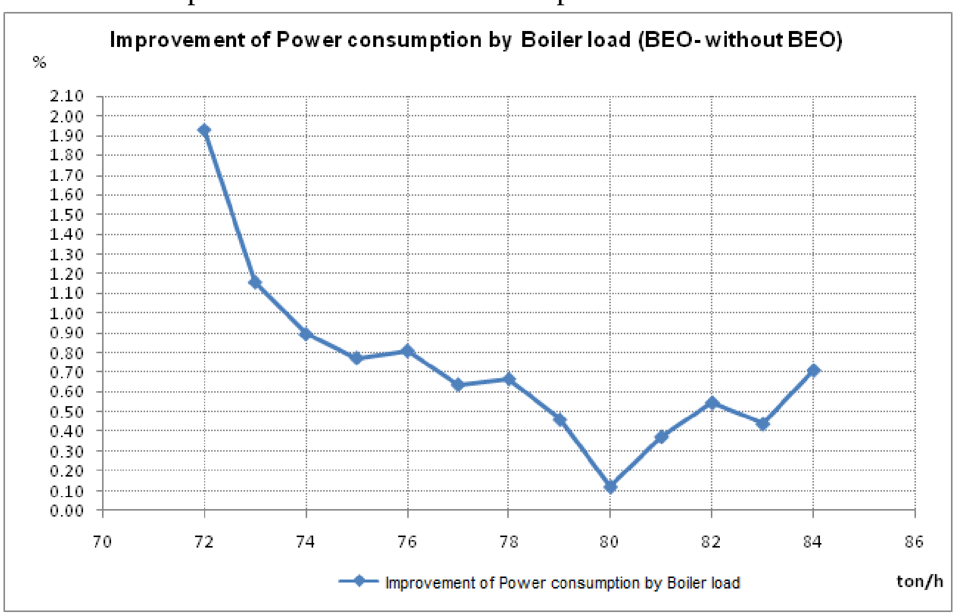

Figure 9: Plot of improvement of Power consumption from 08-11-2013 to 18-11-2013

Estimated benefits by year of applying BEO

Total steam production of boiler at Phu My Fertilizer Plant is approximately 600,000 tons in 2013. The cost of the energy to produce a ton of steam is 2.75 MMBTU/T in average (according to the statistical data of the factory). The average of benefit for 0.52% improved boiler efficiency can be explained as follows:

Year	Steam	Energy to produce 1 ton of steam without BEO	Improved boiler efficiency with BEO	Energy to produce 1 ton of steam with BEO	Energy to produce 1 ton of steam with BEO	Natural gas prices	Benefit to produce 1 ton of steam	Estimated profit for one year of operation
	T/h	MMBTU/T	%	MMBTU/T	MMBTU/T	USD/ MMBTU	USD	USD
2013	1	2,75	0,52%	0,0143	2,7357	6,56	0,09381	56,286.00
2014	1	2,75	0,52%	0,0143	2,7357	6,69	0,09567	57,402.00

Table 1: Estimated Benefit

In conclusion, in reality, the average natural gas price is 6.56 USD/MMBTU in 2013 and 6.69 USD/MMBTU in 2014, the total benefit estimated is about 57,000.00 USD per year. The benefit achieved proves the information system bring more efficient for the boiler operation.

6. Conclusions

This paper presents an information system that has been developed and deployed in Phu My Fertilizer Plant, Vietnam. The system employs data mining techniques, K-means and Fuzzy C-means in particular, and soft computing, recurrent fuzzy neural network in particular, in order to optimize the boiler efficiency of the plant. Currently the system is running at the factory, the results presented in the experiment results reflect the measure of the system effectiveness in short time. As mentioned above, every year, the factory can save about 57,000 USD. The initial results are encouraging and our system shows that improving the boiler efficiency brings productivity, saves cost and increases profit for the factory. That is the motivations for us to continue researching to improve the soft sensor, as well as our system in the future. The most important problem is how to identify exactly when the efficiency of the boiler has any signals of going down. Because the efficiency of the boiler changes continually in small duration so the algorithm recognizing the varying must be very sensitive. Our current solution sometimes does not detect and skip the signals. Our proposed solution in the next time is trying to build a predicting engine that can predict trend of the efficiency of the boiler basing on the current status of the boiler. Artificial neural network or other time series heuristic algorithms may be chosen. In the current, we are researching a pattern matching algorithm and try to prove the effective of this method in theoretical and practical.

References

[1] Nguyen, T. M., Duong, H.N. (2010), “Boiler efficiency optimization by clustering and neural network techniques: data mining approach” The international conference of VietNam Petrol.

[2] Duong, H.N. Tran, Q.M., Dang, K.T. (2009), “Extendable Library of Basic Data Mining Algorithms” The 11st Conference on Science & Technology, Ho Chi Minh City University of Technology, Vietnam.

[3] Emerson Process Management (2009), DeltaV Difference brochure, http://www2.emersonprocess.com/siteadmincenter/PM%20DeltaV%20Documents/Brochures/DeltaV-Difference-Easy-Brochure.pdf

[4] Emerson Process Management (March 2009), DeltaV Neural Product Data Sheet, http://www2.emersonprocess.com/siteadmincenter/PM%20DeltaV%20Documents/ProductDataSheets/PDS_DeltaV_Neural.pdf
[5] Japan Cooperation Center, Petroleum (February 2008), TR-20-07 Recent Sort sensors, Japan.
[6] Aspen technology (2008), AspenONE® Advanced Process Control for chemicals, http://www.aspentech.com/brochures/aspenone_advanced_process_control_chemicals.pdf
[7] Energy Efficiency Guide for Industry in Asia (2006) – Thermal Energy Equipment: Boilers & Thermic Fluid Heaters, www.energyefficiencyasia.org
[8] Duong, H.N. (2006), "Using RFNN to Predict Price of Products in Market" The 9st Conference on Science & Technology, Ho Chi Minh City University Of Technology, VietNam.
[9] Emerson Process Management (2004), Advanced Process Control Solutions, www.EmersonProcess.com/solutions/AAT.
[10] Li, J. Q., Liu, J. Z., Niu, Y. G., & Niu, C. L. (2004), "Online self-optimizing control of coal-fired boiler combustion system" In TENCON 2004. 2004 IEEE Region 10 Conference (Vol. 500, pp. 589-592). IEEE.
[11] Kusiak, A. (2002), "A data mining approach for generation of control signatures" Transactions-American Society of Mechanical Engineers Journal of Manufacturing Science and Engineering, 124(4), 923-926.
[12] Burns, R. (2001). Advanced control engineering. Butterworth-Heinemann.
[13] Kwan, S. (2008), "Kn3 – Power Plant Optimization and Other Applications" OC User Conference-Kuala Lumpur.
[14] ASME (1998), PTC-4 Performance Test Code, Fired Steam Generating Units. USA.
[15] Wikipedia, http://en.wikipedia.org/wiki/Boiler
[16] Montgomery, D.C. (2009), Introduction to Statistical Auality control, Sixth Edition, John Wiley & Son, Inc.
[17] Cogent Real Time System, http://www.opcdatahub.com/WhatIsOPC.html.