

Mining Temporal Data: A Coal-Fired Boiler Case Study

Andrew Kusiak and Alex Burns

Intelligent Systems Laboratory, Industrial Engineering
3131 Seamans Center, The University of Iowa
Iowa City, IA 52242 – 1527, USA
andrew-kusiak@uiowa.edu

Abstract. This paper presents an approach to control pluggage of a coal-fired boiler. The proposed approach involves statistics, data partitioning, parameter reduction, and data mining. The proposed approach was tested on a 750 MW commercial coal-fired boiler affected with a fouling problem that leads to boiler pluggage that causes unscheduled shutdowns. The rare-event detection approach presented in the paper identified several critical time-based data segments that are indicative of the ash pluggage.

1 Introduction

The ability to predict and avoid rare events in time series data is a challenge that could be addressed by data mining approaches. Difficulties arise from the fact that often a significant volume of data describes normal conditions and only a small amount of data may be available for rare events. This problem is further exacerbated by the fact that traditional data mining does not account for the time dependency of the temporal data. The approach presented in this paper overcomes these concerns by defining time windows.

The approach presented in this paper is based on the two main concepts. The first is that the decision-tree data-mining algorithm captures the subtle parameter relationships that cause the rare event to occur [1]. The second concept is that partitioning the data using time windows provides the ability to capture and describe sequences of events that may cause the rare failure.

2 Event Detection Procedure

In the case study discussed in the next section rare events can be detected by applying the five step procedure. These five steps include:

Step 1: Parameter Categorization

The parameter list is divided into two categories, response parameters and impact parameters. Response parameters are those that change values due to a rare event or a failure, e.g., an air leak in a pressurized chamber.

Impact parameters are defined as parameters that are either directly or indirectly controllable and may cause the rare event. These are the parameters that are of greatest interest for the determination of rare events.

Step 2: Time Segmentation

Time segmentation deals with partitioning and labeling the data into time windows (TWs). A time window is defined as a set of observations in chronological order that describe a specified amount of continuous observations. This step allows the data mining algorithms to account for the temporal nature of the data. The most effective method to segment the data is by determining/estimating the approximate date of failure and set that as the last observation of the final time window.

Step 3: Statistical and Visual Analysis

This step involves statistical analysis of the data in each time period that was designated in the previous step. Process shifts, changes in variation, and mean shifts in parameters are helpful in indicating that the appropriate time windows and parameters were selected.

Step 4: Knowledge Extraction

Data mining algorithms discover relationships among parameters and an outcome in the form of IF ... THEN rules and other constructs (e.g., decision tables) [1], [5]. Data mining is natural extension of more traditional tools such as neural networks, multivariable algorithms, or traditional statistics. In the detection of rare events, the decision-tree and rule-induction algorithms are explored for two significant reasons. First, the algorithms generate explicit knowledge in the form understandable by a user. The user is able to understand the extracted knowledge, assess its usefulness, and learn new and interesting concepts. Secondly, the data mining algorithms have been shown to produce highly accurate knowledge in many domains.

Step 5: Analysis of Knowledge and Validation

This step deals with validation of the knowledge generated by the data mining algorithm. If a validation data set is available it should be used to validate the accuracy of the rules. If no similar data is available then unused data from the analysis or a 10-fold cross-validation can be utilized [6].

3 Power Boiler Case Study

The approach proposed in this research was applied to power plant data. Data mining algorithms are well suited for electric power applications that produce hundreds of data points at any time instance.

This case study deals with an ash fouling condition that causes boiler shutdowns several times a year on a commercial 750 MW tangentially-fired coal boiler. The ash fouling causes a build up of material and pluggage in the reheater section of the boiler. Once the build up becomes substantial the boiler performance is negatively affected. This leads to the derating and the eventual shutdown of the boiler. The cleaning of the boiler during the shutdown requires 1 to 3 days. This problem is made more difficult by the fact there is no method to determine the level of ash build up without shutting down the boiler to physically inspect the area. Furthermore, in analysis all parameters were within specifications, so there was no obvious single parameter that is causing the pluggage.

To investigate the problem considered in this paper, data was collected on 173 different boiler parameters. This included flows, pressures, temperatures, controls, demands, and so on. The data was collected in one-minute intervals over the course of three months. The data collection began directly following a shutdown where the reheater section of the boiler had no pluggage. The collection period ended approximately three months later when the boiler had to be shutdown for pluggage removal. This data set contained over 168,000 observations.

The list of 173 parameters, which included both response and impact parameters, was analyzed. The list was reduced to include twenty-six impact parameters. This parameter categorization and reduction was accomplished with the assistance of domain experts as well as statistical analysis such as correlation and multivariate analysis.

The initial step for time segmenting the data was to determine an approximate date for the failure event. In this application the failure event was defined by the date when the boiler was derated due to the pluggage. The cause of the shutdown was confirmed through visual inspection of the affected region. This date was then set to be the last day of the final time window (TW6).

The windows were set to be approximately one week long. A week was chosen for several reasons. First, the boiler was inspected approximately one month prior to its derating. During the inspection the reheater section of the boiler was completely free of ash. This information provided the knowledge that the pluggage required less than one month to manifest itself to the point of shutdown. It was hypothesized that the pluggage requires several days to build up. Based on this information one week was deemed to be an adequate time window. One week also provided a sufficient number of observations (over 10,000 per window) for the data mining algorithms.

Using the derate date and a one-week-long time window, the data was divided into six time windows shown in Figure 1. Time window 1 (TW1) was included to ensure that there was adequate data to describe normal operating conditions.

There appears to be a process shift between time windows 3 and 5 in Figures 1. The west tilt demonstrates a mean shift during window three and the hot reheat steam temperature displays a mean shift as well as a large increase in variation starting in time window four and culminating in window five. The results of this analysis lead to the hypothesis that the events that lead to the eventual pluggage occur between time windows three and five. It also confirms the selection of parameters and window size.

The data mining approach was then applied to the data set to predict the predefined time windows (decision parameter). The algorithm produced a set of rules that described the parameter relationships in each time window.

The knowledge extracted by the algorithm had an overall 10-fold classification accuracy of 99.7%. The confusion matrix (absolute classification accuracy matrix) is shown in

Figure 2. The matrix displays the actual values and the values predicted by the rules during the cross-validation process.

It can be seen from the data in Figure 2 that there are few predicted values that are off by more than one time window from the actual window. The results provided in the confusion matrix provide a high confidence in the proposed solution approach.

Another test data set was extracted from the week following time window 1 and was labeled time window 2 (Test TW2). The last portion of the data (Test TW3) was obtained from the week after the generator was derated and the outcome was labeled time window 6 (TW6). The total test set contained over 30,000 observations.

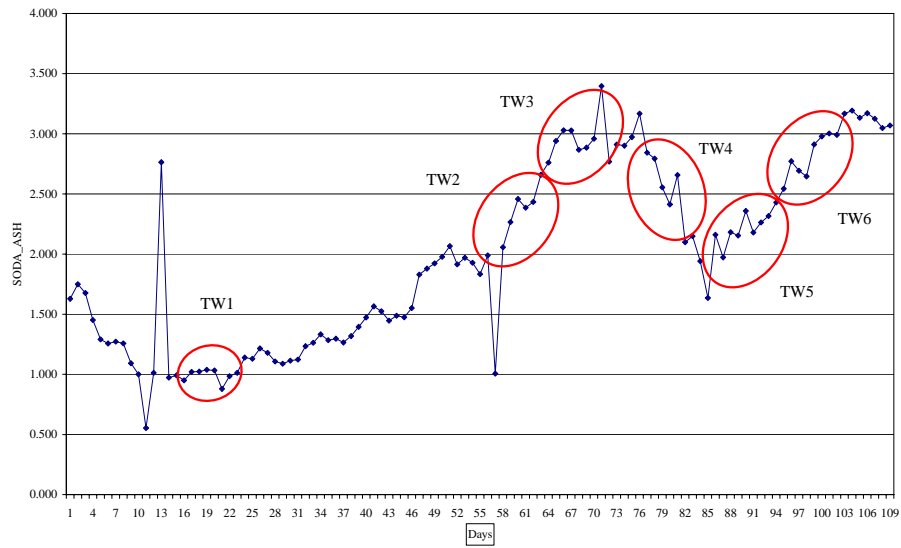


Figure 1. Time windows for ash fouling application.

		Predicted Value					
		TW1	TW2	TW3	TW4	TW5	TW6
Actual Value	TW1	11513	7				
	TW2	17	11497	5			1
	TW3	11		11483	9	5	12
	TW4			13	11461	14	15
	TW5			2	15	12913	30
	TW6	1		1	3	49	12906

Figure 2. Confusion matrix.

The rules and knowledge that were extracted from the original data set were then tested using the test data set. For purposes of analysis time windows 1 – 3 were considered normal and time window 4 – 6 were considered faulty. The resulting confusion matrix is shown in Figure 3.

		Predicted Value		Accuracy
		Normal	Fault	
Actual Value	Normal	19630	370	98.15%
	Fault	2683	7305	73.14%

Figure 3. Cross-validation results for the test data set.

The rules accurately predicted the normal cases, but they were not as effective in predicting the fault cases. This is most likely explained by the fact that the test data labeled, time window 6, was extracted after the boiler had been derated. The derating of the boiler significantly changes the combustion process and was not included in the original data set. In spite of this, the overall classification accuracy of the test data set is greater than 89%. The high cross-validation accuracy indicates that the rules accurately capture the changes in the process that lead to the ash fouling, pluggage, derating, and eventual shutdown of the boiler.

4 Future Research

Event detection for control advisory systems has also been successfully demonstrated for applications that are dynamic and involve rare and catastrophic events [4]. Finch *et al.* [2] developed expert diagnostic information system, MIDAS, to alert users to abnormal transient conditions in chemical, refinery, and utility systems [3].

The approach presented in this research produced rule sets that can be utilized for the development of a meta-control system. Integrating concepts from expert advisory systems and intelligent power control systems will form the meta-control system architecture for the avoidance of the ash pluggage.

5 Conclusion

In this paper a data mining approach to predict failures was proposed and successfully implemented. The research utilized parameter categorization and time segmentation to overcome the limitation of traditional data mining approaches applied to temporal data. The proposed approach produced a knowledge base (rule set) that accurately described the subtle process shifts and parameter relationships that eventually may lead to the detection and avoidance of failures.

This approach was applied to a commercial tangentially-fired coal-boiler to detect and avoid an ash fouling pluggage that eventually leads to boiler shutdown. The approach produced a rule set that was over 99.7% accurate. The knowledge base was also validated with a separate test data set that has predicted failures with accuracy of over 89.8%.

The discovered knowledge will be used to develop an advance warning system reducing the number of boiler shutdowns. The intelligent warning system will have a significant economic impact. This translates into reduced cost to the consumer and a more efficient power industry.

References

1. Quinlan, J.R., "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, 1986.
2. Branagan, L. A., and Wasserman, P. D., "Introductory use of probabilistic neural networks for spike detection from an on-line vibration diagnostic system", *Intelligent Engineering Systems Through Artificial Neural Networks*, vol. 2, pp. 719-724, 1992.
3. Finch, F. E., Oyeleye, O. O., and Kramer, M. A., "Robust event-oriented methodology for diagnosis of dynamic process systems", *Computers & Chemical Engineering*, vol.14, no. 12, pp. 1379-1396, Dec, 1990.
4. Pomeroy, B. D., Spang, H. A., and Dausch, M.E., "Event-based architecture for diagnosis in control advisory systems", *Artificial Intelligence in Engineering*, vol. 5, no. 4, pp. 174-181, Oct, 1990.
5. Pawlak, Z., *Rough Sets: Theoretical Aspects of Reasoning About Data*, Boston: Kluwer, 1991.
6. Stone, M. "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society*, vol. 36, pp.111-147, 1974.