APPLICATION OF FUZZY CLUSTER MEANS FOR DECISION MAKING IN INDUSTRIAL PROCESS

Luís Caldas, lcaldas@fei.edu.br
Fabrizio Leonardi, fabrizio@fei.edu.br
Patrícia Belfiore, prspbelfiore@fei.edu.br
Centro Universitário da FEI
Av. Humberto de Alencar Castelo Branco, 3972, Assunção, CEP: 09850-901, São Bernardo do Campo, SP, Brazil.
Abstract. *This article proposes the use of data mining for evaluating the performance of a cold-rolling plate aiming to infer the maintenance of the equipment by the team. The procedure takes as input a database with samples collected from a lamination process, using multivariate data analysis to assess the dynamics of the system drive of the cylinder mill. Several factors, such as the level of oil in the hydraulic cylinders, may alter the dynamics of the plant, causing loss of efficiency in the lamination of the driver automatically. However, these direct measurements are impossible. The purpose of this study is to show that the position of the cluster centers identify these changes efficiently. The method used is the FCM - fuzzy cluster means. To illustrate the proposed procedure, we adapt a mill model to represent the plant and data, collected via simulation.*

Keywords*: cluster, data mining, decision making, process control.*

1. INTRODUCTION

The market demand to improve quality in manufacturing processes has increased considerably in recent years. When a failure in the product is detected, the image of the company related to their suppliers is affected, causing huge losses. The company needs to invest in technologies linked to its production process and equipments. Besides, the company needs to know what is produced and how it is produced.

The quality of the final product is not compromised by the requirements listed, but by the malfunction of components that don't respect the technical requirements and cause the variability in the production process.

This scenario occurs in the production processes and involves the mills that are metal-processing equipments whose purpose is to reduce the thickness of the raw material from each pass until the desired thickness. This reduction is achieved through cylinders of work driven by hydraulic actuators which are mounted in order to control the opening of the material passage in the process known as gap.

The performance index is linked to the accuracy in the opening and closing of cylinders that operate normally, compensating the variation of the raw material thickness by the variation of the lamination strength.

Several internal and external factors can compromise the quality of the actuator response and consequently they can change the lamination strength, resulting in a non-compliance problem in the final thickness. One of the main factors that impact the lamination strength is the change in the oil level of the hydraulic cylinder because of internal leaks (SIEMENS, 1976, 1979, 1998). This variation of the oil level in the actuator is reflected as a reduction in the course of work in the actuator and, consequently, in the lamination strength. The result is the production of defective materials in the final thicknesses.

This article proposes the use of multivariate data analysis to assess the dynamics of the drive system of the mill's cylinder and therefore helping in decision making for the interference in the equipment by the maintenance team.

The input data composes a database with samples collected from the lamination process. The position of the groups' centers or clusters identifies these changes efficiently. The method of grouping used is the fuzzy cluster means. To illustrate the proposed procedure, a model of a mill available in the literature is adapted to represent the plant and data, collected via simulation.

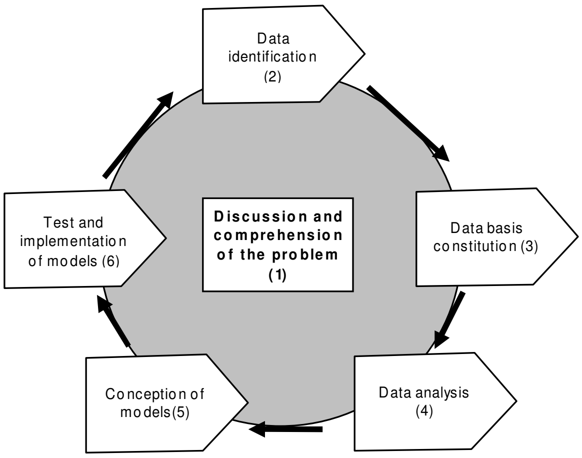
This article is organized as follows. Section 2 gives a brief introduction about data mining and cluster analysis. Section 3 presents the methodology proposed that is based on the data collected and in the application of cluster method. Section 4 describes dynamic systems. Section 5 shows the sampling data, while section 6 presents and discusses the results. Finally, the conclusions are listed in section 7.

2. DATA MINING

According to Bishop (1998), the origins of data mining were in the 60's and it was based on traditional statistics. The conventional statistic is based on small samples and uses behaved and static data. The data mining was originated from conventional statistical techniques as regression. Over the years, a set of techniques, such as fuzzy logic, neural networks, decision trees and other techniques of artificial intelligence were added to the traditional statistics. The data mining has the following characteristics: large volume of data, dynamic, inconsistent and not behaved data.

The data mining can be defined as the application of statistical techniques and artificial intelligence in large amounts of data, with the goal of discovering relationships, trends and relevant patterns among the data (Fayyad, 1996).

Almeida et al. (2005) define the process of data mining in six steps, as shown in Figure 1.



Source: Almeida et al. (2005)

Fig. 1. The process of data mining

According to the authors, the process of data mining begins with the definition of the problem, including purpose and actions for decision making. The second step (data identification) must determine the variables, the amount of data, location and cost of extracting the data to be collected.

The data should be stored in an appropriate manner, through the data warehouse (DW), which is the third step. Statistical analysis of data (fourth stage) is the quantification of information, estimation and inference in order to reduce the uncertainty of the data. This step will begin the knowledge exploitation process. The fifth and sixth stages are the functions and purposes of data mining which are: description and visualization, classification, prediction and grouping.

According to Passari (2003), forecasting aims to provide future behavior. The description and visualization aim to increase the understanding of a complex phenomenon, using techniques of descriptive statistics and tools for displaying graphics. The classification aims to identify the characteristics of an object and put them in a series of pre-defined classes. Finally, the segmentation groups homogeneous clusters of a heterogeneous population. This technique is also called clusters analysis.

The cluster analysis is one of the data mining techniques. Besides the clusters analysis, there are techniques of regression trees, neural networks, fuzzy logic, among others. In this paper, we use the fuzzy clustering technique that uses the concepts of cluster analysis and fuzzy logic.

Fuzzy logic is used to capture vague information, usually in a natural way, and convert them into a digital form in order to facilitate the analysis. The technique is based on the theory of fuzzy sets that, in many cases, are reaching superior results when compared to conventional statistical (BISPO, 1998).

The clusters analysis classifies objects or individuals in relation to some selection criterion pre-determined. The resulting clusters of objects should show high internal consistency within the groups and high heterogeneity between groups. Thus, if the classification is successful, the objects within the groups are close to each other, and the objects of different groups are distant from each other (HAIR et al., 1998).

The method is also an important exploratory technique that seeks to identify a structure of groups aiming to assess the dimensionality of the data, identify outliers and provide interesting hypotheses about associations (Johnson and Wichern, 1992, p. 573).

There are different methods of cluster analysis: non-hierarchical and hierarchical method. In the hierarchical method, each element is isolated in a cluster. Then, there are aggregations of clusters until they get a single cluster with all elements. In the non-hierarchical method it must be defined, a priori, the number of clusters. In general, it is used a sub-sample and a hierarchical method to obtain an estimate of this number. The non-hierarchical method most used is the *k*-means method (average of the attributes of objects).

According to Alvares (2003), the *k*-means method requires the prior definition of the number of cluster and the positioning of the center of each cluster in the area of attributes. The center of the cluster is called centroid. This algorithm is sensitive to noise, but in terms of performance is relatively efficient for large databases.

Alvares (2003) defines the steps of *k*-means algorithm:

Step 1: Select the objects that are centers of the initial *k* clusters;

Step 2: Connect each object to one cluster;

Step 3: Recalculate the centers of the clusters according to the attributes of all objects belonging to the cluster.

Step 4: Return to step 2 until the centers of the clusters stabilize.

According to the author, at each interaction, the objects are grouped according to the center of the closest cluster and, consequently, the centers of the clusters are revalued (step 3), resulting in the displacement of average centers. The algorithm is interrupted when the averages are no longer displaced, or there is a small reallocation of objects between the clusters.

The aim of this paper is to implement multivariate statistical techniques based on fuzzy cluster means algorithm to control industrial processes with multiple input and output variables (lamination process, for example). The purpose of this technique is:

a) Identify the variations in the drive system of work cylinders because of factors with malfunction;
b) Quantify the performance of the mill.

The fuzzy clusters means (FCM) is an algorithm which group data clustered in a specified level according to the degree of relevance. Bezdek proposed this algorithm in 1973. The FCM algorithm divides a collection of n vectors x_i ($i = 1, 2, ..., n$) in c fuzzy groups, and finds the center of each cluster in each group by minimizing the objective function by a dissimilarity measure.

The FCM employs the fuzzy partitioning where a point can belong to several clusters with different degrees of relevance. To represent the fuzzy partitioning, it is used a matrix U of relevance whose elements are in the range [0, 1]. It is reasonable to assume that points in the middle region between clusters have a degree of relevance in both clusters. The relevance of the total points of all clusters, however, should be equal to 1 to keep the properties of the matrix U.

The fuzzy clusters means algorithm (JANG, 1997) determines the centers of the clusters c_i and the matrix U of relevance using the following steps:

Step 1: Start the matrix U of relevance with random values between 0 and 1 respecting the restriction;
Step 2: Calculate the fuzzy center c of clusters c_i (i = 1, 2,..., c);
Step 3: Determine the objective function: stop the process if the result is below a certain level or below a certain amount of tolerance;
Step 4: Determine the new matrix U;
Step 5: Go back to step 2 if the centers of the clusters did not change positions.

3. PROPOSED METHODOLOGY

The methodology used for evaluation the performance of the mill extract patterns from a database. The forms of knowledge acquisition of a database can be accomplished in several ways. As the mill is a dynamic system, we should incorporate the time dependence of the system response. This can be done through predictors applied to the states of the model.

The database is composed of a fixed input excitation and the response of the hydraulic actuator. The database is formed by a series of one hundred items collected for each variable during the process of rolling and at each step the data are read by devices which measure the final thickness. The sensors that are fully allocated in the equipment belong to the automatic control system of thickness, known as AGC. Besides these sensors send the data to control, they still are continuously transmitted to a data acquisition system that makes storage.

These input and output data, collected during the process, are standard to be applied to the cluster analysis technique. Below, the grouping tool is applied and the results are the positions of the clusters' centers that represent the partition of space. The final work for the technical maintenance is to evaluate if any change occurred in the dynamics of the plant by the displacement of the centers.

To verify the effectiveness of the tool and evaluate the results, two sets of data were collected from the mill: the first set of data represents the mill operating in normal conditions and the second represents the mill subject to factors that influence the dynamic response of the system.

The mill used for evaluation is a reversible cold rolling mill, being two work cylinders and two cylinders to support, besides a hydraulic actuator to control the opening and closing of the gap between the cylinders. The dynamics of the mill is represented by a second-order dynamic system and a database collected from a numerical simulation of the model. The signal *wxe* represents the input signal $u(k)$ applied to the mill model.

The same excitement of entry was maintained in all the tests for analysis and evaluation of performance, and parameters of the dynamic model of the hydraulic actuator of the mill were varied.

Figure 2a shows the response of the drive system of rolling cylinders of $x(k)$ *e* $x(k+1)$ and Figure 2b shows a scatter with $x(k)$, $x(k+1)$ and an input $u(k)$.

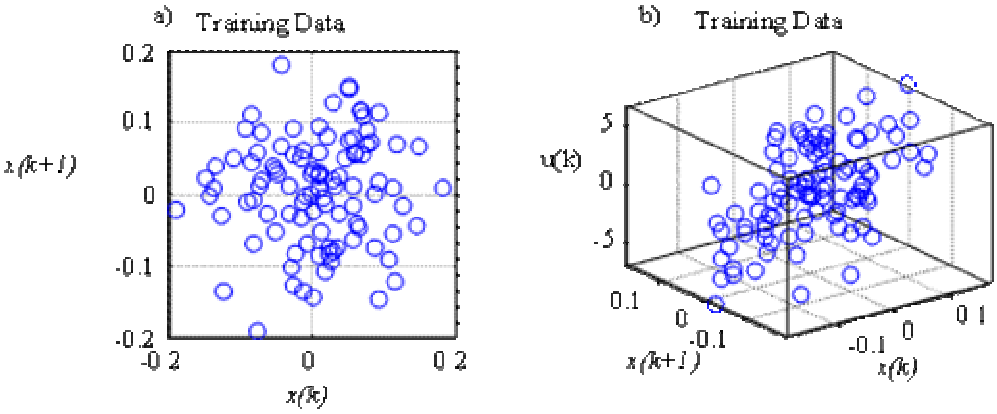


Fig. 2a. Distribution of data collected over time; Fig.2b. Scatter with *u(k), x(k) e x(k+1)*

4. DYNAMIC SYSTEM

Although the system in question is dynamic, the sensitivity of the system response to the changing of the plant may be small, making unfeasible the inspection of the data time of input and output for the identification of failures. Considering the input, it was employed a Gaussian noise with zero mean and unity standard deviation, so that all the dynamic modes are excited because the dynamics of the system may not be known. We used a 2[nd] order linear dynamic model (Franklin, 2002) to represent the system, whose transfer function is given by:

$$\frac{\omega_n^2}{s^2 + 2\varsigma\omega_n s + \omega_n^2} \tag{1}$$

where ω_n is the undamped natural frequency and ς is the damping coefficient. The graphs of Figure 3 show the data acquired of the system output during normal operation and during another operation which there was a change in the parameters of its dynamic model.

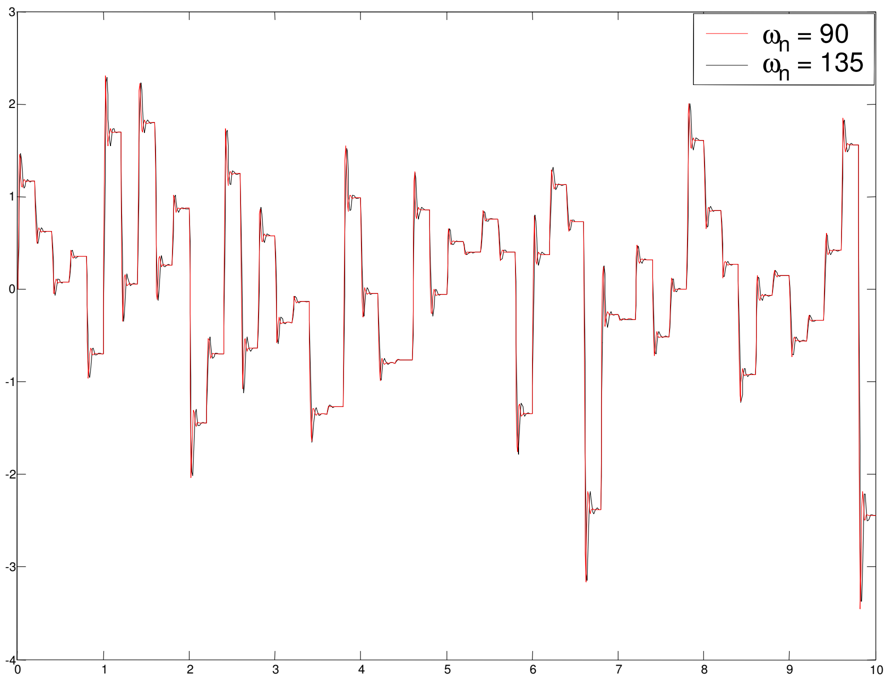


Fig. 3. (a) Answer with ω_n = 90 rad/sec –black (b) Answer with ω_n = 135 rad/sec –red

Note that, although the change of ω_n have been substantial, the dynamic differences are not perceptible by the inspection of the response time. The change from the natural frequency not damped of a system of 2nd order represents a change in resonance frequency of the system that is, in fact, a relevant change. For this reason, it is necessary a more sensitive method to change the parameter. In the next section, we will present the cluster method of data that will be more appropriate.

5. SAMPLING

We collected two types of samples: in the first one the mill operates at normal conditions where its production reaches the levels of tolerance required in terms of quality; in the second the mill operates in non-conformance of some components. For each of the collections, the number of samples used for analysis and treatment were equal to 100, and each sample contained two variables *u(k), x(k)*. The variable *x(k+1)* is the response of the dynamic system which is in the current state *x(k)* and receives an excitement of input *u(k)*, and *x(k+1)* is the future state of *x(k)*. The application for cluster analysis is a toolbox that runs on Matlab software, which uses the algorithm FCM. Setting two clusters and adopting the Euclidean distance, simulations were made with the purpose of determining the centers of each cluster. The two collections were used in the analysis of the clusters and then the results should be compared with each other to evaluate the correlation between the position of the centers and the dynamics of the plant. Figure 4 shows the result of one of the data collection using the cluster analysis by FCM.

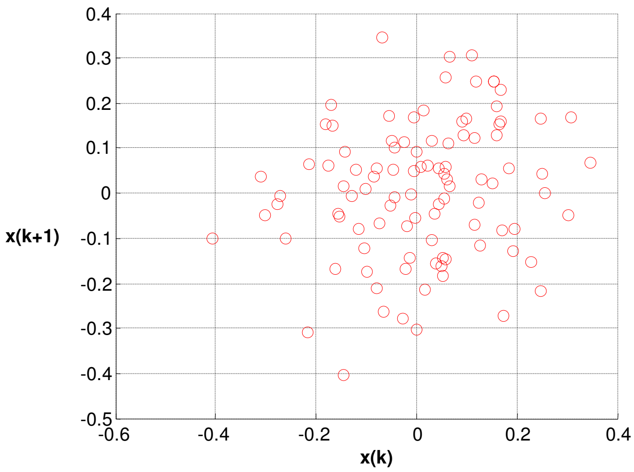


Fig. 4. Sample data collected to analyze the clusters by FCM technique

6. RESULT ANALYSIS

Normally one of the graphs used to evaluate the dynamics of a system is its time domain response. However, this case was not suitable since the output response did not modify substantially, even for a significant change in one of the parameters of the model. An alternative for the dynamic representation of the system is through the phase plan plot. In that plot, the time variable is omitted and only the states are shown. The inconvenience is that for a flat view only two states can be shown in each plot.

Using the phase plan plot it is expected that the change in the dynamics could be noticed by the shape of points. In this work, we proposed that an appropriate way, although heuristics, of doing that by means of the center of the clusters formed in the phase plan. Thus it provides a metric that could be evaluate the change in the dynamic through a single parameter.

The most significant parameter in the analysis was the natural frequency ω_n of the model. The graph of the points collected and the clusters centers for an operation under normal conditions of the hydraulic actuator are shown in Figure 5. The chart for an operation under abnormal conditions of the hydraulic actuator (change in the ω_n value) is shown in Figure 6.

It can be noted by Figures 5 and 6 that the positions centers have changed suggesting a change in dynamic behavior of the plant for the change of natural frequency ω_n at a level sufficient for a decision to stop the equipment by a technical and correct the problem.

The application we have chosen is favorable in the sense that the order of the model is reduced and it allowed to evaluate the dynamic change just for a single phase plan. In the case of more complex dynamics, an analysis more elaborated may be necessary.

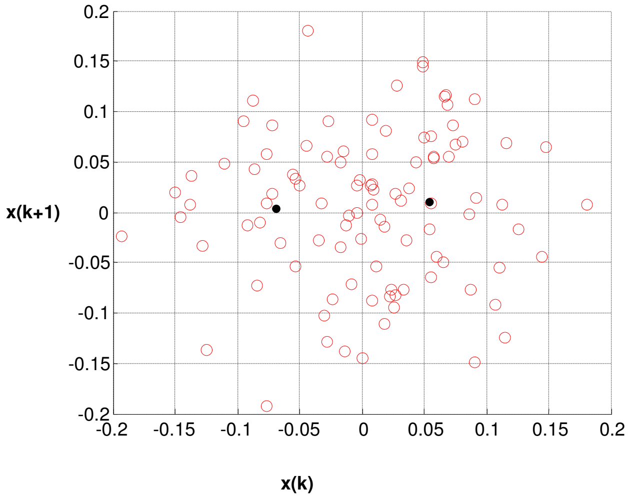


Fig. 5. Distribution of the points of the dynamic system accordingly.

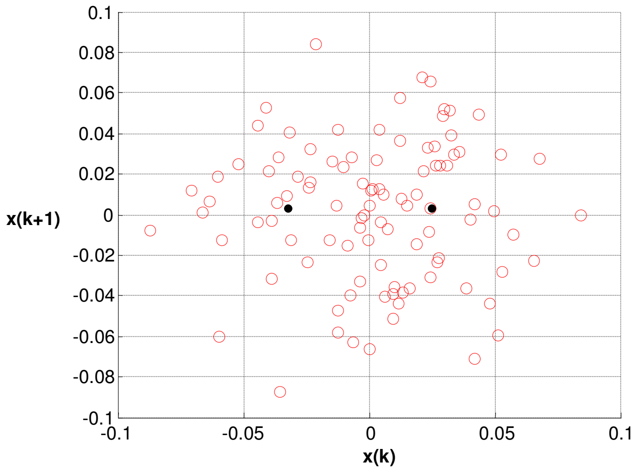


Fig. 6. Distribution of the points of the dynamic system out of line

7. CONCLUSION

This article proposed a solution to a well known problem in the field of lamination, to support decision-making, in order to determine the moment of intervention in the equipment aiming not to impact the production process. Some solutions focus primarily on the question of the oil level, which is one of the factors that can change the dynamics of the plant. It's a bit complicated identify all factors that are almost hidden and significantly influence the dynamic response of system performance.

Many attempts of solutions have been tried successfully and others failed when applied. Some failed technically and others due the operation because the environment is aggressive and may be contaminated with metal particles from the wall or the shirt of hydraulic cylinders.

The solution proposed in this paper is based on the reading of data in the process and then we apply a cluster analysis, allowing infer changes in the dynamics of the plant in a simple and sensitive way. By the graph generated, we can see the level of displacement and the position of the cluster centers, being easy for the technical to make decisions in the moment when he should interfere about the equipment, ensuring the quality of the products through the restoration of the minimum conditions for the operation.

The use of a multivariate statistical technical is one of the steps of data mining, and this tool opens a perspective for further work in the industrial field and supports the statistical process control.

8. REFERENCES

Almeida, F. C., Siqueira, J. O., Onusic, L. M., 2005, "Data Mining no Contexto de Customer Relationship Management", Caderno de Pesquisas em Administração, Vol. 12, No.2, pp.85-97.

Álvares, L.O.C., 2003, "Algoritmos e Ferramentas de Descoberta de Conhecimento em Bancos de Dados Geográficos", Dissertation, Universidade Federal do Rio Grande do Sul, Porto Alegre.

Bezdek, J.C., 1973, "Fuzzy Mathematics in Pattern Classification", Thesis, Applied Math. Center, Cornell University, Ithaca.

Bispo, C.A.F., 1998, "Uma Análise da Nova Geração de Sistemas de Apoio à Decisão", Dissertation, University of São Paulo, São Carlos.

Fayyad, U., Piatetsky-Shapiro, G., Smith, P., 1996, "From Data Mining to Knowledge Discovery: an overview", In: Advance in Knowledge Discovery & Data Mining, pp.1-34.

Franklin, G.F., Powell, J.D., Emami-Naeini, A., 2002, "Feedback Control of Dynamic Systems", 3rd Ed. Addison-Wesley.

Hair, J.F., Anderson, R.E., Tatham, R. L., Black, W. C., 1998, "Multivariate Data Analysis", 5 ed. New Jersey: Prentice Hall.

Jang, J.S.R., Sun, C.T., Mizutani, E., 1997, "Neuro Fuzzy and Soft-Computing", Prentice Hall.

Johnson, R., Wichern, D., 1992, "Applied Multivariate Statistical Analysis", 3 ed. New Jersey: Prentice Hall.

Mathworks Inc, 1996, MATLAB Control System Toolbox – User's Guide.

Passari, A.F.L., 2003, "Exploração de Dados Atomizados para Previsão de Vendas no Varejo utilizando Redes Neurais", Dissertation, University of Sao Paulo.

Siemens, A.G., 1976, "Industrial and Building Group, Cold Rolling Mills, Processing Lines", Electrical Equipment and Automation for single stand mills – Solutions for Metals, Mining and More.

Siemens, A.G, 1979, "Industrial and Building Group, Cold Rolling Mills, Processing Lines", Electrical Equipment and Automation for Cold Strip Tandem Mills – Solutions for Metals Mining and More.

Siemens, A.G., 1998, "Industrial and Building Group, Cold Rolling Mills, Processing Lines", New Thickness Control Mode Based on Mass Flow Principle Increases Cold Rolling Accuracy – Ideas for Steel.

9. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.