Application of a neural network to classify the out-of-control signal that gives the T2 multivariate graph of Hotelling using data obtained in the industry

Gerardo Avendano. PhD
Faculty of Business Administration, EAN University
Bogota Colombia
gavendanop@universidadean.edu.co

Gustavo Andrés Campos-Avendano PhD
Faculty of Mathematics and Engineering
Fundación Universitaria Konrad Lorenz, Bogotá, Colombia
gustavo.campos@konradlorenz.edu.co

Abstract

In the development of industrial processes there are situations where it is necessary to control or simultaneously monitor two or more quality variables of the production process. The problems of process monitoring where several related variables are studied can be controlled by means of multivariate control charts. The objective of this work is to describe the implementation of the control chart Hotelling's T2 using real data obtained from the industry.
In the present research work the implementation of the control chart Hotelling's T2 in the industrial field is carried out and a real multivariate case is analyzed with quality variables.

Keywords
Key words: Statistical control of multivariate processes, graphic Hotelling's T2.

1. Introduction

The simultaneous control of several variables in the same piece or product has received much attention in recent years. The use of a single multivariate graph for process control is a more powerful option than using several univariate graphs. However, multivariate graphs do not indicate which variables or variables are being measured that have changed and current techniques are inefficient to detect the variables responsible for the change. Therefore the research will be focused on finding a method based on neural networks in order to identify the variables that have changed in the process, using industrial data Montgomery (2009).
The main objective of this research is to design a generalized method based on neural networks to classify the out of control signal that the Hotelling T2 multivariate graph Aparisi (2014), applied to a production process with real data obtained from the industry.
The simultaneous control of several variables in the same piece or product has received much attention in recent years. The use of a single multivariate graph for process control is a more powerful option than using several univariate graphs. However, multivariate graphs do not indicate which variables or variables are being measured that have changed and current techniques are inefficient to detect the variables responsible for the change. Therefore, the research will be focused on finding a method based on neural networks in order to identify the variables that have changed in the process, using industrial data.

2. Approach and formulation of the research problem

Identify the variables responsible for the change in an industrial process monitored by means of a multivariate control chart. With the advent of computers in the eighties, the neural networks took great strength since they overcame the obstacle that for years prevented them from developing, the low computational capacity of computers. This is how a large number of applications arise for neural networks; one of them is the classification and in our particular case the classification of the signal out of control, responsible for the change in the multivariate control charts.
The production processes and the quality control that are currently applied in the companies allow to detect changes in the production process that imply a loss of quality in the product. Today quality control charts are a widely used tool in the industry, Campos (2012).

The simultaneous control of several quality variables in the same piece or product has received a lot of attention in recent years. If we want to control p variables simultaneously, we have two approaches: Use *p* univariate graphics. For example, *p* type graphics, CUSUM, or EWMA. Use a single multivariate graph. The options in this case include, among others, the Hotelling's T2 chart, the generalized variance graph |S|, Aparisi (2004), or the various options in the MCUSUM chart.
The use of multivariate graphs proves to be a more powerful option, that is, it requires, on average, fewer points in the graph to detect a change in the process compared to the use of univariate graphs, Lowry and Montgomery (1995). This is a cost savings for the company, since in the final production you get a better quality in the product. It seems then that the multivariate option would clearly be the choice Khoo (2004).

3. Methodology

- Bibliographic review of existing methods to solve the proposed problem.
- Verification of the effectiveness of these methods, measured as the percentage of hits in the case of analytical methods, and the relevance of application in the case of graphical methods. Like the analysis of the application software of neural networks that exist today to find the one of greater relevance in the industrial application proposed.
- Validation of the results obtained through statistical techniques. That is, by means of hypothesis test procedures to verify the best neural network structure found.

4. Process: Optimization of the network for 3 variables

For the training of the network in the case of three variables were used the variables weight, thickness and diameter of the bottles manufactured in a brewing industry.

Using the data from the bottles of the brewing industry, the values of the Hotelling's multivariate T2 control chart were obtained through the Mathcad program and after several tests the best network structure for three variables was found.

The optimal network found is a network of 4 layers in total: 8 nodes in the input layer, 14 in the first inner layer, 10 in the second inner layer and 3 nodes in the output layer. In this case the sample size, the type of point on the graph, the distance of Mahalanobis and the cut-off point were varied.

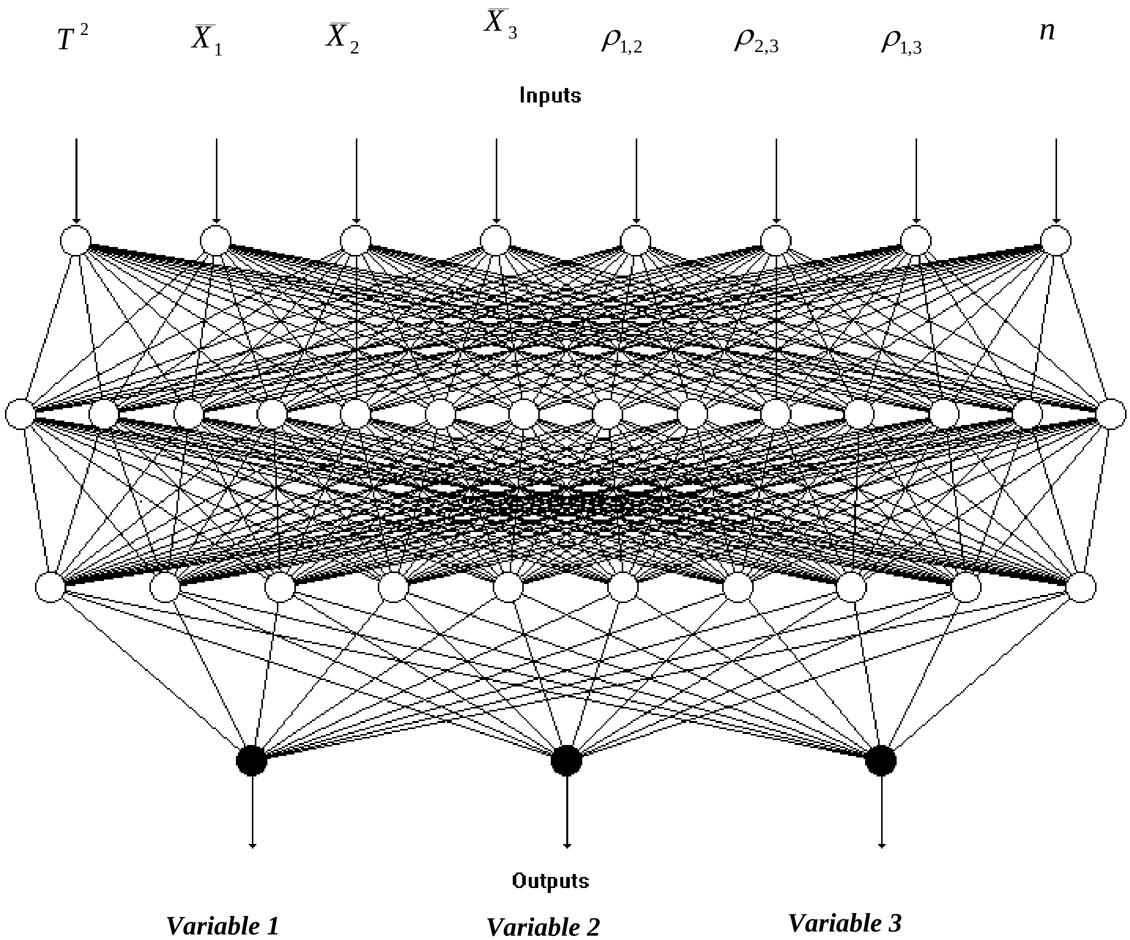

Figure 1 Structure of the neural network used, case p = 3.

4.1 Statistical significance

In order to analyze the results obtained in the success percentages of the neural networks with three values of the T2 statistic, we developed a hypothesis test that analyzes the differences between the networks Grasman (2010).

Hypothesis test on the population proportion L of a binomial distribution

L1 Percentage of network hits with a single point (69.5%)
L2 Percentage of net hits with previous point (71.3%)

Where the null hypothesis is H0: (L1 = L2) y The alternative hypothesis is: H1: (L1 L2)

We Accept H_0 si $\left|\bar{L}_1 - \bar{L}_2\right| \leq Z_{\alpha/2}\sqrt{L^* q^* (1/n_1 + 1/n_2)}$

We reject H_0 si $\left|\bar{L}_1 - \bar{L}_2\right| > Z_{\alpha/2}\sqrt{L^* q^* (1/n_1 + 1/n_2)}$

$$\text{With } L^* = \frac{n_1\hat{L}_1 + n_2\hat{L}_2}{n_1 + n_2} \quad \text{y} \quad q^* = 1 - L^*$$

When performing the calculations, a p-value of 0.5 (greater than 0.05) is obtained, which is why we accept the hypothesis of equal success percentages in the two types of neural networks. That is, there is no significant difference in the results found in the percentage of hits that detect neural networks with one or two points. Thus, it does not seem reasonable to use a much more complicated neural network that produces the same percentage of hits as the previously designed network.

4.2 An Example

Applying the Hotelling's T2 control chart to the data in table 1 obtained from the bottle production process in the brewing industry we have:

Table 1. Bottle production data

Sample	Weight Container	Thickness center	Body Diameter
1	14,98	2,01	52,35
2	14,99	2,14	52,32
3	14,97	2,14	52,25
4	15	2,24	52,23
5	14,99	2,12	52,23
6	15	2,1	52,23
7	14,96	1,91	52,23
8	14,93	2,11	52,28
9	14,96	2,02	52,28
10	14,99	2,21	52,39
11	14,97	2,19	52,24
12	14,98	2,09	52,32
13	15,01	2,06	52,31
14	14,99	2,04	52,3
15	14,91	1,97	52,3
16	14,98	1,91	52,24
17	14,95	1,94	52,27
18	14,97	2,09	52,27
19	14,98	2,11	52,33
20	14,99	2,12	52,43
21	14,95	2,05	52,42
22	14,99	2,02	52,3
23	15	2,08	52,31
24	14,96	2,11	52,23
25	14,98	2,07	52,29

Statistic to be plotted Hotelling (1947):

$$T_i^2 = n(\vec{X}_i - \vec{\mu}_0)' \sum_0^{-1} (\vec{X}_i - \vec{\mu}_0)$$

LC Control Limit: 10.37

Phase 1 - estimated covariance from the current data using changed differences.

Table 2. T2 Hotelling chart with a value outside the LC.

Gráfico	Alfa	LIC	LSC	Fuera de Límites
T-Cuadrada	0,005	0,0	10,3712	1

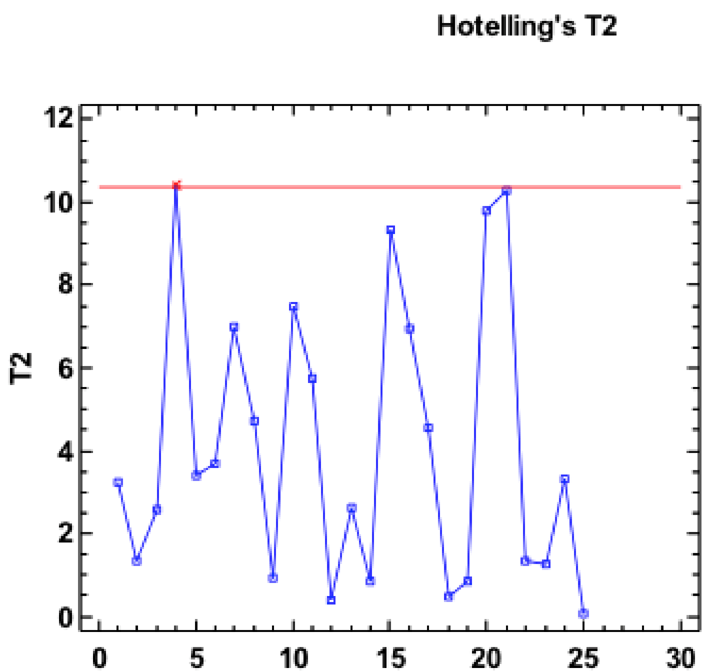

Figure 2. T2 Hotelling chart using three process variables in the bottles changed.

Example of the variables of the bottles:

Table 3. Dimensions for the application example using the T2 Hotelling chart, n =1

Bottles	Weight Container	Thickness center	Body Diameter
Sample 4	15.00	2.24	52,23

The value of T2 that has been observed applying Hotelling's T2, is 10.4276 which is a value greater than LC 10.3712. That is, the process is not under control. As we see the T2 multivariate graph of Hotelling detects the signal of lack of control but does not say that variable is responsible for the change, for this we will use the neural network.

CONCLUSIONS

As seen in the above analysis, Hotelling's T2 chart detects the out-of-control signal but does not say which one or which of the variables are responsible for the change or failure in the process. For this we use the software developed by Aparaisi. F Avendaño G. Sanz J. (2006) Obtaining the following results

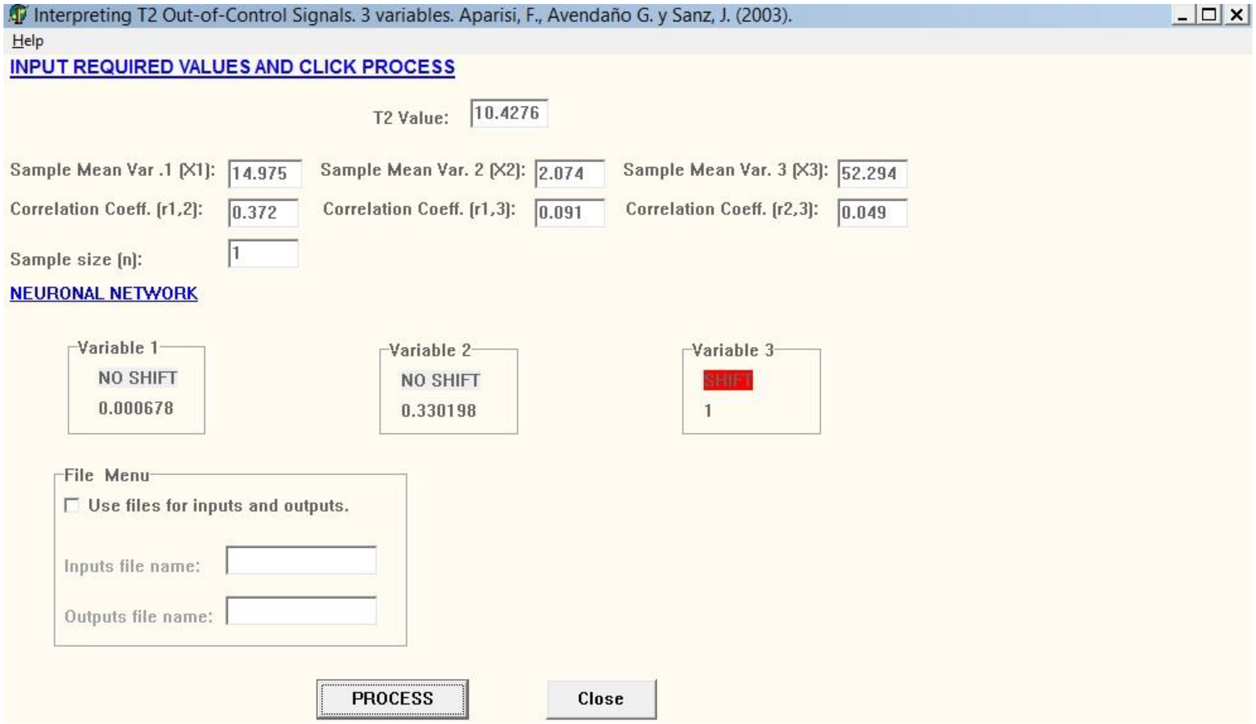

Figure 3. Software output screen using Hotelling's T2 chart.

The program shows that the change produced by the process, and detected by the Hotelling's T2 multivariate control graph, corresponds to the third variable. Which in this example corresponds to the diameter of the bottle.

It is concluded that the network found in the investigations of Aparisi. F., Avendaño. G., Sanz. J. (2006) has been used. Which has the ability to find out which variables have changed in the process when the Hotelling T2 chart indicates that a change has occurred.
The designed network is a general network, in the sense that:

- It serves any productive process, whatever the variables are measured.
- Detects all possible types of changes.
- We can change the sample size, since the sample size is an input to the network.

Increasing the number of variables increases the difficulty for the network to detect the variables that indicate the state of out of control of the process, since the number of possible combinations between variables increases (only one variable, groups of two variables, groups of three Variables and the four variables).

For the practical use of this procedure has been used the software developed by Aparisi F. Avendaño. G., Sanz. J. (2005) program that is simple to manage, in which the end user does not need to have knowledge of neural networks. That is, the user must enter the input data of the network in the corresponding box, and the program responds as the variable that has changed.

References

Aparisi, F., Avendaño, G., & Sanz, J., Techniques to interpret T2 control chart signals. IIE Transactions, vol. 38, no 8, p. 647-657, 2006.

Aparisi, F., Epprecht, E., Carrión, A., & Ruiz, O. The variable sample size variable dimension T 2 control chart. *International Journal of Production Research*, vol. 52, no 2, p. 368-383, 2014.

Aparisi, F.; García-Díaz, C., Optimization of univariate and multivariate exponentially weighted moving-average control charts using genetic algorithms. *Computers & Operations Research*, vol. 31, no 9, p. 1437-1454, 2004.

Campos, G., Caracterización y optimización de los gráficos X-RL2 para el control de la posición y S-RL2 para el control de la dispersión. Tesis Doctoral. Tesis Doctoral, Universitat Politècnica de València, 2012.

Grasman, R. P., Huizenga, H. M., & Geurts, H. M. Departure from normality in multivariate normative comparison: The Cramér alternative for Hotelling's T2. *Neuropsychologia*, vol. 48, no 5, p. 1510-1516, 2010.

Hotteling, H. Multivariate Quality Control in Techniques of statistical Analysis, *edited by Eisenhart, Hastay and Walls*. 1947.

Khoo, M. B. C.; Quah, S. H.; Low, H. C., Powerful Rules for the Hotelling's $\chi 2$ Control Chart. *Quality Engineering*, vol. 17, no 1, p. 139-149, 2004.

Montgomery, D. C. Statistical quality control. New York: *Wiley*, 2009.

Biography

Gerardo Avendano is a full-time professor at the EAN University in Bogotá Colombia with more than 20 years of experience in teaching and research and management at prestigious universities and national and international companies. Experience in the industry in the area of production and quality, as well as in personnel management and group work. Author of scientific publications such as books, articles in indexed journals and software development. Postdoctor in Statistical Process Control SPC Penn State University (State College, USA 2007). Ph.D. Polytechnic University of Valencia (Valencia - Spain, 2004). Specialist in Advanced Statistical Methods for the Improvement of Productivity and Quality Polytechnic University of Valencia (Valencia Spain 2001). Specialist in Production Engineering of the District Francisco José de Caldas Bogotá Colombia (1998). Chemical Engineer National University of Colombia (1994).

Gustavo Andrés Campos-Avendano is a fulltime Professor, and Director of Research of Faculty of Mathematics and Engineering in Fundación Universitaria Konrad Lorenz, Bogotá, Colombia. He is PhD in statistics and optimization from Universidad Politecnica de Valencia, Spain. He has published books, journal and conference papers.