Data analysis techniques for neural networks-based virtual sensors

Thomás López-Molina*, Anna Pérez-Méndez* Francklin Rivas-Echeverría,**
Universidad de Los Andes
*Escuela de Estadística
**Laboratorio de Sistemas Inteligentes
Mérida, Venezuela 5101

Abstract:- This work presents the use of statistical techniques for data analysis and processing for its use in the resolution of the most common problems of greater importance in the development of virtual sensors using artificial neural networks. The potentialities of use of such statistical techniques are studied and an example of an industrial process application appears.

Key Words:- Data Analysis Techniques, Artificial Neural Networks, Virtual Sensors, Statistical analysis.

1 Introduction

Artificial intelligence [1, 2, 8, 9, 19, 44] is one of the scientific areas with greater diffusion and application in the last years. Every day is more common to find industrial, commercial or academic tools that involve the use of intelligent techniques in the resolution of critical and recurrent problems. The neural networks [9, 19, 22, 30] could be considered as one of the most spread and used artificial intelligence techniques due to their simplicity, implantation facilities and design characteristics.

Virtual sensors are a virtual instrumentation area whose main objective is to generate estimations of important variables using the information about past values of the variable or other process variables measurements that are related to the one desired to be predicted. Neural networks have been one of the most used intelligent tool for designing and developing Virtual Sensors due to them accurate, modelling and Identification capabilities and easy for implantation [37].

The neural networks use; greater every day in different human knowledge areas, has created additional requirements and necessities of data processing in training systems. Users are interested, among other things, in having tools that allow them to filter and depurate the input data, to select the most relevant variables and training patterns and to complete missing values.

On the other hand, the statistical data analysis techniques [3, 13, 28, 29, 31, 32] have been applied to an increasing number of knowledge areas in recent years. They are particularly appropriate for the study of great data volumes in which it is impossible, due to its size, to observe structural immediately characteristic.

This work treats on the use of such techniques in order to solve the main problems found in artificial neural networks training systems and is structured as follows: Section 2 presents the suggested techniques for the input variables and patterns selection to neural networks. Section 3 contains the techniques for atypical values detection. Section 4 treats on virtual sensors and neural networks. Section 5 presents by means of an example the data analyses techniques evaluation and finally, section 6 contains the pertinent conclusions and recommendations.

2 Variables and Patterns Selection

In this section several techniques appear for diminishing the number of input variables to Neural Networks training systems. First, they are selected according to their prediction capacity for the output variables (Elementary Analysis). Then, linear combinations are made in order to obtain new variables that try to summarize the information provided by the original ones (Principal Components Analysis). Additionally, strategies for making data partitions in training and validation sets are depicted: the Simple Random Sampling and the Stratified Random Sampling, which provide the partition size and indicate how the patrons must be selected.

2.1 Elementary analysis

The Elementary Analysis is an statistical technique that selects the input variables according to its capacity for explaining the output variables. It obtains an "improvement" measurement when incorporating variables and then it takes decisions on the best set determined for a particular size. It was proposed originally by Newton and Spurrell [33]; Hintze [21] presents a generalization of the technique for the multivariate case in which the relation is given between a set of outputs variables and another set of inputs. It uses the Wilk Criterion [43] for calculating the measures that are used for obtaining the conclusions.

2.2 Principal Components Analysis for reduction of Variables Number

The Principal Components Analysis [3, 23, 25, 35] is one of the Statistical Multivariate Analysis classic techniques. It is used with no need of distributional assumptions for describing a data set from certain geometric and algebraic optimized criteria [24, 26, 27, 28]. Next, it is presented the main criteria that have been established for reducing the data dimensionality using Principal Component.

Selection Strategies using Principal Components

Next, the criteria for the components selection in the data dimensionality reduction that could be useful for the construction of Neural Networks are presented:

- Eigenvalues must be greater than certain established value
- Proportion of Explained Variance by the Components must be greater than certain established value
- Component whose Eigenvalues are highly correlated with the output variables
- Elementary Analysis
- Combined Strategy of high Eigenvalues and more correlated low ones [11].

2.3 Patterns Selection

A commonly used procedure by neural networks users consists of splitting the data in two excluding sets, one for training and another one for validation. Usually there become empirical partitions as for example 70-30 (70% for training and 30% for validation). The accomplishment of these partitions has evident limitations: I. They do not take into account the dispersion of input data (variables with high dispersion require more data for training), II. They do not take into account the multidimensionality of the data (different dispersions can suggest different partitions), III. The partitions can produce bad generalization in the neural network if the patterns are not selected suitably.

In this section two useful techniques are depicted that will help to select the partition size by means of a well stated statistical criterion and additionally will make the selection of training and validation patterns: Simple Random Sampling and Stratified Random Sampling [7, 34].

When a neural networks user has a set of N patterns and wishes to make a partition in n_e training and n_v validation patterns ($n_e + n_v = N$), he can randomly select one of the $\frac{N!}{n_e!(N-n_e)!}$ possible training sets, such selection is known as Simple Random Sampling [7]. This sampling is going to provide the partition size (n_e or n_v) and will indicate which patterns must be including in each one of the partitions, according to randomness and representativeness.

The patterns selection is made one by one: They are enumerated from 1 to N, a series of n_e random numbers is extracted (if the training patterns are going to be selected) in such a way that in each successive extraction the same opportunity of selection to all patterns that have not yet been selected is granted. The selection of the number of patterns to be included in each partition must be related somehow to the error that has been allowed. From the statistical point of view, the partition size or "sample size" is related to the Maximum permissible error (r) and the probability of committing this error (α). This allows obtaining n_e or n_v as a function of the error and of certain parameters considered through the available data. Since generally it is used more data for training than for validation, the size of the partition (n) that is going to be obtained will be n_e if it happens that $n > 50\% N$, in another case it will be n_v.

Let's suppose now that between the data set it is available a classification variable that split the N training data in sub-groups of $N_1, N_2, ..., N_L$ units, that are not overlapped and that as a whole includes the totality, this is:

$$N_1 + N_2 + ... + N_L = N$$

Each sub-group will be denominated "cluster". The Stratified Random Sampling is the procedure which makes an independent Simple Random Sampling within each cluster. This sampling technique provides greater representativeness in the partition in relation to the Simple Random Sampling when there is knowledge about the clusters, when there are noticeable differences of sizes in them or when their variabilities differ. A refinement of this procedure is the Stratified Principal Component Analysis [11].

3 Atypical values detection

Even in the data collected with greater care, some times observations can appear that seemed not to belong to the set. There are two concepts related to this problem: I. The Spurious Observations, Suspicious, Aberrant, Atypical or just “Outliers”, are those that in opinion of the investigator seem not to belong to the original set of data. II. Discordant Observations are Outliers that after being put into the discordance Statistical test [38, 39, 41] were not free from all suspicion with security (1-α) [17, 18].

3.1 Multivariate Outliers Detection using Box-and-Whisker Plots

Univariate or marginal data, obtained of each multivariate data component, can be useful in the Outliers detection. Some reasons are [6]: I. has an appropriate geometric idea of the univariate Outliers: those observations that are suspiciously moved away from the original data set, located in the extremes, II. It is possible that there exist detected multivariate Outliers in some particular components (variable), like severe measurement errors in some of the observed variables. Additionally, the univariate techniques for Outliers detection can be applied to a function of the observed variables, like a linear combination provided by Principal Components [20] and then the Multivariate Outliers can be detected.

One of the most popular graphical techniques for the exploratory data analysis is the Box-And-Whisker Plot [43] that allows, among other things, to detect those suspiciously remote observations of the data.

Tukey [42] have introduced the construction of the Box-and-Whisker Plot as a simultaneously form for observing 5 statistics from the data set: median, lower and upper quartile, minimum and maximum. It defines as step the value of 1,5 times the interquartile rank and indicates that values outside the quartiles. Near under a step are considered “Adjacent”, those that are between one and two steps are considered “Outliers” and beyond two steps are considered as “Far Outliers”.

3.2 Identification using Fourier Series type functions

Geometrically it is only possible to represent points in two or three dimensions spaces and point’s three-dimensional representation in paper is difficult to elaborate and to understand. Andrews [4] have proposed a graphical method for projecting k-dimensional points in a bi-space based on the univariate functions representation. Let X=(x1,x2,...,xk)’ a k-dimensional vector of quantitative variables and defining the function:

$$f_X(t) = \tfrac{1}{\sqrt{2}} x_1 + x_2 Sen(t) + x_3 Cos(t) + x_4 Sen(2t) + x_5 Cos(2t) + ...$$

The function fX(t) is plotted in $[-\pi, \pi]$ and a set of n points will appear by means of n different curved. Andrews [4] have presented an example in which it is demonstrated the utility of this graphical representation for forming groups and Gnanadesikan [14, 15] proposes an adaptation of the plot for studying the configuration and distributional aspects of the multivariate data when n is big. The general concept is as follows: It is selected a set of p values of t in $[-\pi, \pi]$, that is $\{t_0, t_1, t_2, \dots, t_{p-1}\}$, better if they are equally spaced, and for each ti it is located the median, percentage points and particular values of fXi(tj) (Gnanadesikan [15] uses p =100).

In relation to the percentage points selection there is much freedom on the matter. Gnanadesikan [15, 16] suggests the use of 10%, 5% and 1% or probability equally spaced in 12½%, 6¼% and 31/8%; also it is possible to use points moved away between one and two steps of first and third quartile like in the Box-and-Whisker Plots. The lower and upper limits that are more moved away from the Median are used for detecting Outliers. Gnanadesikan additionally suggests to identify Outliers in the Quartile Contour Plot if their curves f are move away (superior or inferiorly) from the limits for a proportion w of ti’s values (for example more than 50%). Naturally, while in more ti’s it is determined an observation as Outlier, it would be more sure that it certainly is an Outlier.

As it is indicated by Barnett and Lewis [5], the Andrews plot [6] and the Gnanadesikan modification [15, 16] they are techniques useful for detecting Outliers simultaneously, being able to be used even with transformed variables (for example, using Principal Component, logarithms, etc). In the curves, the presences of Clusters or Outliers are evidenced by means of those far from the others avoiding the Masking Effect (when an Outlier is hidden in another observation near it). Other uses of these graphs have been proposed by Andrews [4] and Gnanadesikan [16], as control of multicolinearity or normality assumption.

4 Virtual Sensors and neural networks

Virtual sensors constitute an area of virtual instrumentation, whose principal mission is to perform indirect measurements of process important variables using historical data of the desired variable and some other variable that affects its performance [37]. Some authors have defined virtual sensors as: "… the numerical prediction or estimation of a desired system state given other related physical signals such that the output of the virtual sensor estimates the true state in real time, on-line in an open-loop configuration"[12].
Virtual sensors are some times designed for working in parallel with a physical sensor in order to evaluate its performance, but they can be used also for having on-line estimation of the desired measurement.
Virtual sensors are widely used because they are computer programs that can be change or updated when it is necessary. Virtual sensors can be designed using neural networks in order to learn the laboratory tests results or the signals given by a physical sensor (Figure 1).
During the last years, neural networks-based virtual sensors [5, 10, 12] have been widely used in industrial environments because of it potentiality for identifying complex nonlinear dynamical systems and for having appropriate results in different situations.

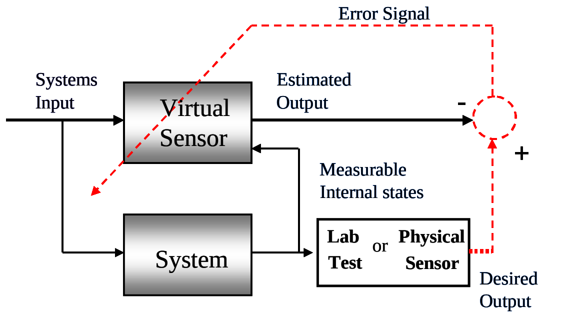

Figure 1. Virtual sensor using the Laboratory test or Physical sensor results

5 Data analysis techniques evaluation for virtual sensors using neural networks

In this section it is presented an industrial application example for the evaluation of the techniques presented in sections 2 and 3. In order to make the performance test, several neural networks were trained considering different data sets obtained using the techniques presented in the previous sections. It was designed a virtual analyzer for estimating the quality of a Houk type gasoline obtained in a catalytic cracking unit [36, 37].
The refinement process is very complex and is divided in sections or units. In the catalytic cracking unit of a refinery plant there are measured more than 180 chemical and physical variables. In this unit it is obtained different products from the crude as: alcohols, gas oils, gases, different kinds of gasoline, etc. The mission of this unit is to break the carbon chains in order to obtain lighter ones, which are more commercially attractive. The Houk gasoline is obtained from the plate 33 of the catalytic cracking unit. The quality of the Houk gasoline may be expressed in terms of it final point temperature (temperature when the total evaporation of the gasoline in a lab test occurs).

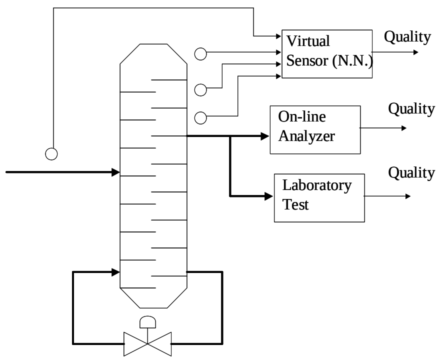

Figure 2. Catalytic Cracking Unit

The work consists in obtaining an estimation of the final point temperature of the Houk gasoline in order to obtain an on-line estimation of the product quality (that is obtained using laboratory analysis only two or three times a day).
The measurements of the processes variables are given by sensors that are located in different parts of the unit, the quality variables are measured using on-line analyzers and there are some other performance variables that are calculated using some computer algorithms. All this data is stored in databases.
In the catalytic cracking unit there are some measured variables as: Temperatures, pressures, flows, and some calculated variables that are related with the desired quality. It has been used on-line analyzers and laboratory test for estimating the gasoline quality or final point temperature. The data

acquisition system collects information about 183 variables. Some studies given by expert process engineers have determined that only 14 of those variables have some influence on the quality level. These 14 variables are quantitative and are measured using sensors or are calculated from other measured variables.

For the design of the virtual sensor, it was constructed a data matrix containing the obtained measurements of the temperature final point and the other important variables, selected by the experts, that can be used for estimating the quality variable. The obtained file contains 15 variables and 3721 observations for each variable. This data set needs to be processed in order to eliminate those patterns (observations) that are out of specifications. Next table presents a briefly description of each of the variables:

	Min.	Max	$\bar{X}$	S^2
PTO Q	28	252.95	227.98	1045.98
TP33	143.42	153.08	148.329	3.06552
TALIM	180.98	190.2	184.662	4.2435
PP33	1.20	1.29	1.25626	0.000223
VCBE	15.69	29.01	19.2361	1.5859
FSAL	214.13	522.36	377.135	3811.72
VCFQ	18.11	26.34	21.4979	1.1222
VCTP	8.15	16.45	13.1993	1.3304
TTOP	93.94	104.73	97.4146	4.503
TP34	132.21	141.95	137.53	3.701
TP32	145.98	155.23	151.24	3.650
TSAL	96.17	100.408	96.682	0.482
FALIM	839.81	1953.4	1511.4	21353.8
TFON	185.75	194.93	189.066	5.1234
PTOP	11.67	12.59	12.222	0.02628

	S	Range	S_{KP}	K
PTO Q	32.3416	224.62	93.521	201.079
TP33	1.751	9.66	-3.3643	-6.45
TALIM	2.0599	9.32	16.6123	-6.862
PP33	0.0149	0.08	-9.9121	-2.4676
VCBE	1.2593	13.32	26.8546	46.5912
FSAL	61.739	308.23	1.589	-13.4581
VCFQ	1.059	8.23	-0.634	-1.3553
VCTP	1.1534	8.3	-7.73	-2.4253
TTOP	2.122	11.79	10.731	-1.7121
TP34	1.924	9.74	-11.891	-2.056
TP32	1.9106	9.25	-9.889	-5.984
TSAL	0.6945	4.31	-16.1022	12.4015
FALIM	146.129	113.59	-6.061	-0.0497
TFON	2.2635	9.18	23.9817	-1.7934
PTOP	0.1621	0.92	-14.6441	-1.885

A briefly description of each of the selected variables:

- Pto. Q: Final Point Temperature. This is the quality variable that is going to be estimated using the virtual sensor.
- TP33: Temperature on Plate 33.
- TALIM: Feeding crude temperature.
- PP33: Pressure on Plate 33.
- VCBE: Adimensional Variable Calculated using the energy balance.
- FSAL: Gasoline recolection flow.
- VCFQ: Adimensional variable Calculated using crude physical-chemical properties.
- VCTP: Adimensional Variable Calculated using Temperature and Pressure.
- TTOP:Temperature on the Top of column.
- TP34: Temperature on the Plate 34.
- TP32: Temperature on Plate 32.
- TSAL: Temperature of the gasoline at the end of the unit.
- FALIM: Feeding crude flow.
- TFON: Temperature at the bottom of the column.
- PTOP: Pressure at the top of the unit.

The complete description of the variables can be appreciated in [36, 37].

The first used technique was the Quartile Contour Plot with Tukey limits for detecting Outliers (selection: 50% of $p = 100$). There were detected 8 Outliers. Found results are shown in table 1.

Table 1. Outlier's location, Outlier Percentage and Far Outlier Percentage

. Pattern Number	%Outliers in Partition	% Far Outliers in Partition	Status
2972	72	0	Outlier
2973	73	0	Outlier
2974	83	6	Outlier
2975	83	9	Outlier
2976	83	8	Outlier
2977	81	1	Outlier
2978	72	0	Outlier
2979	56	0	Outlier

As it is available a great amount of data and the Outliers percentage is very low (inferior to 0,3% of the total) it was decided to eliminate these considered Outliers patterns, nevertheless this must be made with precaution when it is not available a great amount of data, because the elimination can induce to a significant diminution of the effective training sample size.

Once eliminated the Outliers, a Principal Components Analysis was made with the correlations matrix of the 14 variables to be used as inputs. Principal Components 1, 2, 3, 9 and 10 were selected. At this moment we have two data sets with 3713 patterns each one: A. Data with the 15 original variables, B. Data with the Five selected Components.

In order to make the partition of the data for Training and Validation, Simple Random Sampling was applied to the 15 original variables. The results of this technique were not appropriate (ne:3710; nv: 3). The reason of this strange partition is the great dispersion of some variables, needing great samples according to the error and established confidence level. A strategy for diminishing this error is to make data stratification [11]. Applying the Sturges rule, 13 classes were constructed according to the output variable (PTO_Q). The new size under the Stratified Random Sampling partition appears in table 2.

Table 2. Partition Size according to Stratified Random Sampling

Cluster	N	Training	Validation
1	14	13	1
2	12	11	1
3	66	59	7
4	11	10	1
5	11	10	1
6	15	13	2
7	12	11	1
8	4	3	1
9	8	7	1
10	61	55	6
11	80	72	8
12	2179	1958	221
13	1240	1114	126
Total	3713	3336	377

Next, it was performed the neural networks training with the best perceptron networks topologies with two layers suggested by the computational tool STATISTICA Neural Networks 4.0 (Statsoft Inc., 2000) [40] that makes an exhaustive search using Genetic Algorithms [1, 9] to obtain them:

- Training Set A: Input Layer (14 nodes, Linear Function), Hidden Layer (1 node, logistic function) and Output Layer (1 neuron, Linear Function).
- Training Set B: Input Layer (5 nodes, Linear Function), Hidden Layer (4 nodes, logistic function) and Output Layer (1 neuron, Linear Function).

In regression situations as the one in the presented example, the intention of a Neural Network is to learn how to predict the output variable (PTO_Q) from the input variables. Thus, a neural network will be successful if it makes better predictions compared with a simple prediction like the average.

In order to make this comparison the Error Standard Deviation is used, which must be smaller than the Standard Deviation of the output for considering the Neural Network as a good predictor. As it can be appreciated in table 3 both neural networks are good predicting but the obtained using the Principal Components (set B) makes better predictions than the network obtained by means of the gross data (set A), because it presents a smaller Prediction Error Standard deviation for training and validation.

Table 3. Regression statistics for data sets A and B

	Set A		Set B	
	Training	Validation	Training	Validation
Data average	227.3303	228.3405	227.4317	227.5017
Standard deviation	32.35.926	30.41637	32.45516	29.4938
Mean Error	0.359457	-0.3266	-0.3163	1.080847
Error Standard deviation	26.4076	27.18941	25.56667	26.17833
Absolute Mean Error	14.38797	15.28565	12.07537	11.56443
Prediction Error Standard deviation Ratio	0.8160755	0.8939069	0.7877537	0.8875875
Correlation	0.5791918	0.4497229	0.6164413	0.4618089

Another presented performance measurement is the Prediction Error Standard deviation Ratio; when it is smaller it indicates that the neural network is better for making output variable estimations; observe that the neural network obtained with the data set B generates a slightly smaller value of this statistic than the corresponding to set A. Additionally, the Pearson correlation between the present values and the predicted ones appears. Theoretically a correlation of 1,0 indicates a perfect linear relation between the prediction and the present output values, but actually it is a good performance indicator. With this indicator, the neural network trained with Principal Components (set B) as inputs is better than the trained with the Gross Data (set A) in the output variable prediction, as much in training as in validation.

6 Conclusions

The results obtained in this investigation demonstrate the advantages of having statistical analysis to input data previous to training Artificial Neural Networks: Shorter training periods, simpler topologies and more reliable networks can be found. The presented techniques for variables and patterns selection allow reducing the data dimension, obtaining quicker training, simpler topologies and lower prediction errors.

The pattern reduction techniques allow generating a data partition for training and validation based on statistical analysis. Additionally, these selection techniques can be used for reducing the patterns number in the data when it is very high, applying Simple or Stratified Random Sampling and training or validating with the data sets that are obtained using the partition.

The Outliers detection techniques follow the Data Analysis graphical philosophy, so the great volumes of data used for neural networks training make it difficult their immediate use. Nevertheless it is possible to use them for developing algorithms that detect possible observations significantly different from the rest of the data. The neural networks trainer has always the option to use the gross data or depurate and select those that provide a better training.

The Minimal Square technique usually applied in neural networks learning algorithms, is well-known by its lack of robustness to the presence of Outliers, reason why it is necessary to use the possible instruments that allow to detect Outliers in the data that is going to be used in the training phase.

In the designed virtual sensor for quality estimation, there were found improvement using the statistical analysis, because it use generate easier neural network topologies with better performance.

It is very important to continue developing the fusion of both disciplines: Artificial intelligence and Statistical Data Analysis. The works made until the moment presents forts evidences of the advantages that it has for the practical Statistic the Artificial intelligence and vice versa.

7 References

[1] Aguilar, J and Rivas, F. (1999) “Introducción a las Técnicas de Inteligencia Artificial”. Editorial Meritec. Mérida.

[2] Amador-Hidalgo, L. (1997) Inteligencia Artificial y Sistemas Expertos. Servicio de Publicaciones de la Universidad de Córdoba

[3] Anderson, T. (1958). An Introduction to Multivariate Statistical Analysis. John Wiley & Sons Inc., New York

[4] Andrews, D. F. (1972). Plots of high-dimensional data. Biometrics 28, 125-36

[5] Atkinson, C. Traver, M. et al. “Virtual Sensors – A Real Time Neural Network Based Intelligent Performance and Emissions Prediction System for On-Board Diagnostics and Engine Control”. West Virginia University (http://www.cemr.wvu.edu/~virtsens/)

[6] Barnett, V. and Lewis, T. (1978). Outliers in Statistical Data. John Wiley & Sons Inc., New York

[7] Cochran, W. (1977). Sampling Techniques. John Wiley & Sons Inc., New York

[8] Cohen, P. (1989) The Handbook of Artificial Intelligence, New York. Addison Wesley

[9] Colina, E., Rivas, F. (1998) Introducción a la Inteligencia Artificial. Cuadernos de Control. Postgrado en Ingeniería de Control. Universidad de Los Andes

[10] Colina, E. Rivas, F. Rios, A. Chacón, E. Montilva, J. (1997) “Aspectos Metodológicos para la Implantación de Sensores Virtuales”. VI Jornadas de Automatización Industrial de Producción. Venezuela.

[11] Colmenares, G. (1999). Reducing Samples and Variables to Train, Test and Validate Neural Networks. Doctoral Dissertation. Department of Industrial and Management Systems Engineering. College of Engineering. Unversity of South Florida, Florida, EEUU

[12] Deignan, P. (1999). “Virtual Sensing:The Development of a Methodology for Internal Combustion Engine Torque Estimation”. Purdue University.

[13] Freixa, M., Salafranca, L., Ferrer R., Guardia, J. y Turbany, J. (1992) Análisis Exploratorio de Datos. Promociones y Publicaciones Universitarias. Primera Edición. Barcelona España

[14] Gnanadesikan, R. (1973). Graphical methods for informal inference in multivariate data analysis. Bull. Int. Statist. Inst., 45, Book 4, 195-206

[15] Gnanadesikan, R. (1977). Methods for Statistical Data Analysis of Multivariate Observations. John Wiley & Sons Inc., New York

[16] Gnandesikan, R. and Kettering, J.R. (1972). Robust Estimates, Residuales and Outlier

Detection with Multiresponse Data. Biometrics 28, 81-124

[17] González de Q., F. (1977). Test para Múltiples Outliers en Muestras de Poblaciones Normales Univariantes. Trabajo de Ascenso presentado a la Universidad de Los Andes, Facultad de Ciencias Económicas y Sociales. Mérida, Venezuela

[18] González de Q., F. (1984). Comparaciones del Comportamiento de dos Estadísticos para probar Múltiples Observaciones Discordantes. Trabajo de Ascenso presentado a la Universidad de Los Andes, Facultad de Ciencias Económicas y Sociales. Mérida, Venezuela

[19] Hagan, M., Demuth, H., Beale, M. (1996) Neural Networks Design. PWS Publishing Company. U.S.A.

[20] Hawkins, D. (1974). The Detection of Errors in Multivariate Data Using Principal Components. JASA, 69, 346

[21] Hintze, J. (1980). On the Use of “Elemental Analysis” in Multivariate Variable Selection. Technometrics, 22, 4

[22] Honik, K., Stinchcombe M., y White, H. (1989). Multilayer feedforward networks are universal aproximators. Neural Networks, vol 2, 5, pp 359-366

[23] Hotelling, H. (1933). Analysis of a Complex of Statistical Variables into Principal Components. J. Educ. Psychol. 24, 417-441 y 498-520

[24] Johnson, R. A. and Wichern, D. W. (1992). Applied Multivariate Statistical Analysis. Prentice Hall, New Jersey

[25] Jolliffe, I. (1986). Principal Components Analysis. Springer Verlag.

[26] Káiser (1960). The application of electronic computers to factor analysis. Educ. Psychol. Meas. Vol 20, pp 141-151

[27] Kshirsagar, A. (1972). Multivariate Analysis. Marcel Dekker Inc. New York

[28] Lebart, L. Morineau, A. y Warwick, K. (1984). Multivariate Descrptive Statistical Analysis. John Wiley & Sons Inc., New York

[29] Mardia, K., Kent, J., y Bibby, J. (1979). Multivariate Analysis. Academic Press, London

[30] MathWorks (1994). Neural Networks Toolbox for MatLab. MathWorks

[31] Morrison, D. (1990) Multivariate Statistical Methods. Third Edition. McGraw Hill Publishing Company

[32] Neural Works. Neuralware Inc. Website: http://www.teleport.com/~cognizer/eet/Almanac/COMPANY/42088521.HTM

[33] Newton, R. y Spurrel, D. (1967). A Development of Multiple Regression for the Analysis of Routine Data. Applied Statistics, Miscellanea

[34] Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. Jour. Roy. Stat. Soc., 97, 558-606

[35] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Phil. Mag., 2 (sexta serie), 559-572

[36] Pérez, G. (2001). Metodología para el Desarrollo de Sensores Virtuales basados en redes Neuronales. Proyecto de Grado presentado a la Universidad de los Andes, Escuela de Estadística, Facultad de Ciencias Económicas y Sociales, Mérida, Venezuela

[37] Pérez, A. Nava, L. Rivas, F. Colina, E. (2002) “Neural Networks-based Virtual Sensors Methodology”. WSEAS International Conference on Neural Networks. Interlaken, Switzerland.

[38] Shapiro, S. and Wilk, M. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52, 3 and 4

[39] Shapiro,S. Wilk, M. and Chen, H. (1968). A comparative study for various tests for normality. JASA, pp. 1343-1372

[40] Statistica Neural Networks. Website: http://www.statsoft.com/products/stat_nn.html

[41] Tietjen, G. and Moore, H. (1972) Some Grubbs-Type Statistics for the Detection of Several Outliers. Technometrics 14, 3

[42] Tukey, J. W. (1977). Exploratory Data Analysis, Vol 1. Addison-Wesley, California

[43] Wilk, S. (1932). Certain generalizations of the analysis of variance. Biometrika, 39,471-481

[44] Zadeh, F. (1965) Fuzzy Sets. Information and Control pp.338-353