

Proyecto "Optimización de procesos industriales y estudio de series temporales mediante Redes Neuronales Artificiales"

Departamento de Investigación Institucional – Facultad de Química E Ingeniería "Fray Rogelio Bacon"
Pontificia Universidad Católica Argentina – Campus Rosario

Obtención, clasificación y análisis de datos de procesos industriales en empresas del ámbito local con fines de optimización mediante el uso de redes neuronales artificiales

Javier Fornari^a, Eduardo Luccini^{a,b}, Esteban Vidali^a, Miguel Parodi^{a,b}, Sebastián Grieco^a

^a*Facultad de Química e Ingeniería, Pontificia Universidad Católica Argentina, Av. Pellegrini 3314, 2000 Rosario, Argentina.*

^b*Grupo de Energía Solar, Instituto de Física de Rosario (CONICET-UNRosario), Rosario, Argentina*

Resumen. Se identificaron y clasificaron las variables que caracterizan procesos industriales en el ámbito de la industria metalúrgica, como estrategia para modelizar y optimizar el proceso, a partir de los datos de un conjunto de empresas de la región centro y sur de Santa Fe. Como herramienta integradora de análisis se plantea la aplicación de redes neuronales artificiales, en particular mediante mapas auto-organizativos (SOM, Self-Organizing Maps). Los resultados preliminares confirman que el enfoque utilizado es capaz de proporcionar valiosa información y ofrece posibilidades para la aplicación directa sobre la industria local.

Keywords: Procesos Industriales, Redes Neuronales Artificiales, Clasificación, Optimización.

1 INTRODUCCION

En los actuales procesos de fabricación industrial, las plantas están bajo una enorme presión para mantener y mejorar la calidad del producto y el beneficio en estrictas limitaciones medioambientales y de seguridad. En el ámbito de la industria, una pequeña mejora en el conocimiento y la predicción de los procesos involucrados implica un significativo impacto en el resultado económico-productivo de una empresa. La modelización de un sistema de producción con fines de mejorar el proceso se enmarca dentro de la Teoría de Optimización (e.g. Forst y Hoffmann, 2010), disciplina que ha ganado enorme desarrollo en las últimas décadas, aplicada particularmente a la planificación y el estudio del desempeño de líneas de producción industrial (e.g. Rao, 2009). Más generalmente, la simulación numérica permite integrar las distintas etapas de un sistema de producción a fin de optimizar su funcionamiento. Para lograrlo, se deben determinar las variables que rigen el comportamiento del sistema y las leyes que las gobiernan. Los algoritmos y las técnicas de análisis disponibles son muy variados, y su elección depende fuertemente de las características del problema a resolver (e.g. Rao, 2009). En el caso de la industria metalúrgica, algunas de las variables incluyen el tiempo de operación, la secuencia de operaciones, el tamaño del lote de las partes, la tipología del equipamiento utilizado, el nivel de demora de las partes, el scrap o material residual del proceso productivo, entre otras. Entre otros aspectos, la optimización implica el estudio de eventos críticos, o interrupciones en el sistema de producción por diversas causas, y sus consecuencias. Para ello se han desarrollado diversos modelos que simulan las etapas de la línea de producción y los efectos de un evento crítico (Fornari et al., 2010).

Uno de los algoritmos de análisis que ha cobrado más desarrollo y número de aplicaciones en las últimas décadas es el de redes neuronales artificiales (RNA). Una RNA es una arquitectura de procesamiento en paralelo distribuido, integrada por unidades de procesamiento masivamente interconectadas, también denominadas neuronas o nodos. Los modelos más populares de RNA pueden clasificarse basándose en el paradigma de aprendizaje, su topología de conexión básica y sus funciones de procesamiento y capacidades, destacándose las asociadores de patrones o memorias heteroasociativas, las redes competitivas o mapas de auto-organizativos, los modelos de satisfacción de demanda o de adaptación probabilística y otras redes asociativas de pesos fijos. Las RNA han demostrado su potencial en control, optimización y detección de errores en procesos industriales (Zarenezhad y Aminian, 2011; Govindhasamy et al., 2005; Alves and Nascimento, 2007).

En el presente trabajo se detalla la obtención de datos de producción a partir del relevamiento de 21 empresas del rubro metalmecánico de la ciudad de Rosario y 22 empresas de la ciudad de Rafaela en la provincia de Santa Fe, Argentina, con el propósito de analizarlos a fin de optimizar los procesos productivos aplicando la técnica de RNA.

2 OBTENCION DE DATOS

Los datos de procesos industriales se obtuvieron mediante la realización de un cuestionario común para todas las empresas, completada a modo de encuesta, dividida en siete secciones, con un total de 66 preguntas que procuran cubrir las diversas áreas de la empresa para determinar el impacto de las tecnologías en los procesos productivos, cuyo detalle se observa en la tabla 1.

Tabla 1. Características del cuestionario realizado en cada empresa relevada, cuyos datos constituyen la base de análisis del presente trabajo.

SECCION	CONTENIDOS	NUMERO DE PREGUNTAS
A-Información de la empresa	Identificación de la empresa a través de la razón social, persona de contacto y producto o servicio principal, etc.	19
B-Gerencia y Administración	Referentes a la tercerización de diversas áreas o servicios de la empresa, nivel de exportación y uso de indicadores de rendimiento	7
C-Operaciones	Referentes al tipo de insumo que utiliza en su proceso industrial, estrategia de fabricación utilizada (diseño, fabricación y ensamblado bajo pedido o para inventario), tipo y frecuencia de mantenimiento del equipamiento industrial utilizado	16
D-Diseño e Ingeniería de Procesos	Oferta de productos o servicios, mejoras realizadas en los últimos 3 años, ciclo comercial del producto o servicio. Las mejoras están focalizadas en el diseño, especialmente si fue solicitado por el cliente, y en los procesos industriales de manufactura	9
E-Perfil Innovador	Detección del tipo de perfil de la empresa, así como del empresario y su área ejecutiva, visión innovadora de futuro de la empresa y el mercado, incorporación de tecnologías de la información y comunicación, nivel de educación del personal y capacitaciones realizadas en el último año	6
F-Nivel de adopción de TICs	Introducción de tecnologías en los últimos 3 años, la criticidad de las tecnologías en el proceso industrial y el nivel de impacto en su incorporación	7
G-Nivel de utilización de TICs	Tecnologías utilizadas actualmente en la empresa en cualquier sector o área junto con la alineación estratégica de las tecnologías en el proceso industrial y de comercialización de sus productos o servicios	2

La figura 1 muestra como ejemplo una de las páginas de la encuesta, correspondiente al cuestionario de la sección C-Operaciones.

Sección C - Operaciones

28. ¿Qué porcentaje de las ventas tienen como activos en almacenes? (Marque una opción por fila)

	100-75%	75-50%	50-25%	25-0%
Materia Prima y Empaque	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Insumos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Maquinarias y Repuestos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Productos Terminados	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

29. Para las siguientes afirmaciones asigne un puntaje (de 0 a 5) en relación con su grado de conformidad (siendo 0=falso y 5=verdadero)

Realizamos controles de inventario mensualmente	<input type="text"/>
Los pasillos de los almacenes generalmente están limpios y libre de obstáculos	<input type="text"/>
El abastecimiento se realiza desde más de 3 (tres) proveedores por ítem	<input type="text"/>
No tenemos problemas frecuentes de distribución de espacio	<input type="text"/>

30. ¿Alguno de sus almacenes se localiza a más de 500 metros entre sí o del centro de producción?

SI NO

31. Indique cuáles son las estrategias de fabricación predominantes en la empresa

	SI	NO
Diseño bajo pedido	<input type="checkbox"/>	<input type="checkbox"/>
Fabricación bajo pedido	<input type="checkbox"/>	<input type="checkbox"/>
Fabricación para inventario	<input type="checkbox"/>	<input type="checkbox"/>
Ensamblado bajo pedido	<input type="checkbox"/>	<input type="checkbox"/>

32. ¿Considera usted alguno de sus ítems de inventario como estratégicos?
(Si la respuesta es "NO" pase a la pregunta 34)

SI NO

Figura 1. Ejemplo del cuestionario utilizado, correspondiente a preguntas de la sección C-Operaciones.

La figura 2 muestra un ejemplo de la grilla de respuestas obtenidas dentro de la misma sección.

Sección C - Operaciones							
28	MP y Empaque	25	25	25	25	25	25
	Insumos	25	25	25	25	25	25
	Maquinaria y R	25	25	25	25	25	25
	Productos Term	25	25	25	50	50	50
29	Opción A	2	2	2	3	2	2
	Opción B	1	1	1	1	2	2
	Opción C	3	3	3	2	2	2
	Opción D	4	3	3	3	4	3
30		SI	SI	SI	NO	NO	NO
31	Opción A	NO	NO	NO	NO	NO	NO
	Opción B	SI	SI	SI	SI	SI	SI
	Opción C	SI	SI	SI	NO	NO	NO
	Opción D	SI	SI	NO	NO	NO	SI
32		SI	SI	SI	SI	SI	SI
33	Opción A	50%	50%	50%	50%	25%	25%
	Opción B	50%			50%		50%
	Opción C		25%	25%			
	Opción D		25%	25%		75%	25%
34	Opción 0-5	X		X	X	X	
	Opción 6-15		X				X
	Opción 16-30						
	Opción > 30						
35	Opción A						
	Opción B						X
	Opción C						
	Opción D	X	X	X	X		
	Opción E					X	
36		2T	2T	2T	2T	2T	2T
37		SI	SI	SI	SI	SI	SI
38	Opción A						
	Opción B						
	Opción C						
	Opción D	X	X	X	X	X	X
39	A	4	4	5	4	4	3
	B	4	4	5	3	4	5
	C	2	2	1	2	2	1
	D	3	3	3	3	2	3
	E	3	3	3	3	3	3
	F	4	4	3	4	4	3

Figura 2. Ejemplo de la grilla de respuestas a la encuesta realizada en la sección C-Operaciones.

3 PROCESAMIENTO DE DATOS

En función de los datos relevados, y dado que las respuestas eran con opciones y de tipo cerradas, se debió categorizarlas a fin de transformar las respuestas cualitativas en valores cuantitativos. La matriz final está compuesta de 44 filas, donde la primera fila corresponde a los nombres de los atributos utilizados, y 159 columnas que representan las diversas respuestas a las preguntas formuladas. En este caso, existen preguntas que tienen varias secciones y del total de 66 preguntas junto con sus varias secciones determinan el total de columnas mencionado. La figura 3 presenta una grilla resumida de los datos relevados en cada empresa.

Empresa	14	15	16	17	18	19	20	21	22	23	24
1	SA	NO	NO	35	MAS 100	MAS 15	V-L-H-S	SI	20%	5%	NO
2	SA	NO	NO	22	MAS 100	MAS 15	V-L-H-S	SI	15%	5%	SI
3	SA	SI	SI	15	80	MAS 15	V-L-H-S	NO	20%	15%	SI
4	SRL	NO	NO	20	40	10 M	HIG. Y SEG.	SI	20%	5%	NO
5	SRL	NO	NO	10	20	10 M	HIG. Y SEG.	NO	35%	15%	NO
6	SRL	NO	NO	12	20	10 M	HIG. Y SEG.	NO	20%	5%	NO
7	SA	NO	NO	8	80	15 M	V-L-H-S	SI	30%	15%	SI
8	SA	NO	NO	9	80	10 M	V-L-H-S	SI	20%	15%	NO
9	SA	NO	NO	5	60	MAS DE 5	V-L-H-S	SI	15%	5%	NO
10	SH	NO	NO	4	20	MAS DE 15	HIG. Y SEG.	NO	10%	5%	NO
11	SA	NO	NO	7	Mas 100	Mas 15	V - L - H - S	NO	15%	5%	NO
12	SA	SI	SI	6	Mas 100	Mas 15	V - L - H - S	NO	10%	5%	SI
13	SA	NO	NO	8	60	Mas 15	V - L - H - S	SI	5%	5%	NO
14	SA	NO	NO	8	60	Mas 15	V - L - H - S	SI	20%	5%	NO
15	SA	NO	NO	3	60	Mas 15	V - L - H - S	NO	25%	5%	NO
16	SRL	NO	NO	4	20	10 M	Hig y Seg	NO	15%	5%	NO
17	SRL	NO	NO	2	20	5 M	Hig y Seg	NO	15%	5%	NO
18	Monol	NO	NO	1	20	5 M	Ninguno	NO	25%	5%	NO
19	SRL	NO	NO	12	40	10 M	Hig y Seg	NO	15%	5%	NO
20	SRL	NO	NO	8	40	10 M	V - L - H - S	NO	10%	5%	NO
21	SA	SI	NO	16	60	Mas 15	V - L - H - S	SI	20%	5%	NO

Figura 3. Matriz parcial de datos relevados por empresa.

Para proceder a la clasificación de las diferentes empresas relevadas y con el fin de obtener una segmentación que permita representar el estado de situación de tecnología y procesos industriales en forma agrupada, se procedió al uso del algoritmo de clasificación de los mapas auto-organizativos utilizando diversas topologías que son evaluadas para seleccionar el que mejor represente la situación actual de las empresas en estudio. En este caso se considera el mejor adaptado al que cumpla con una serie de condiciones, tales como menor error en la caracterización del problema, velocidad de respuesta y complejidad del modelo para su implementación, entre otros criterios. La validación del modelo seleccionado se realizó por medio del software de aplicación que implementa los diferentes modelos de RNA, en este caso el producto Matlab® versión 7. La figura 4 muestra los aspectos más representativos de las RNA implementadas.

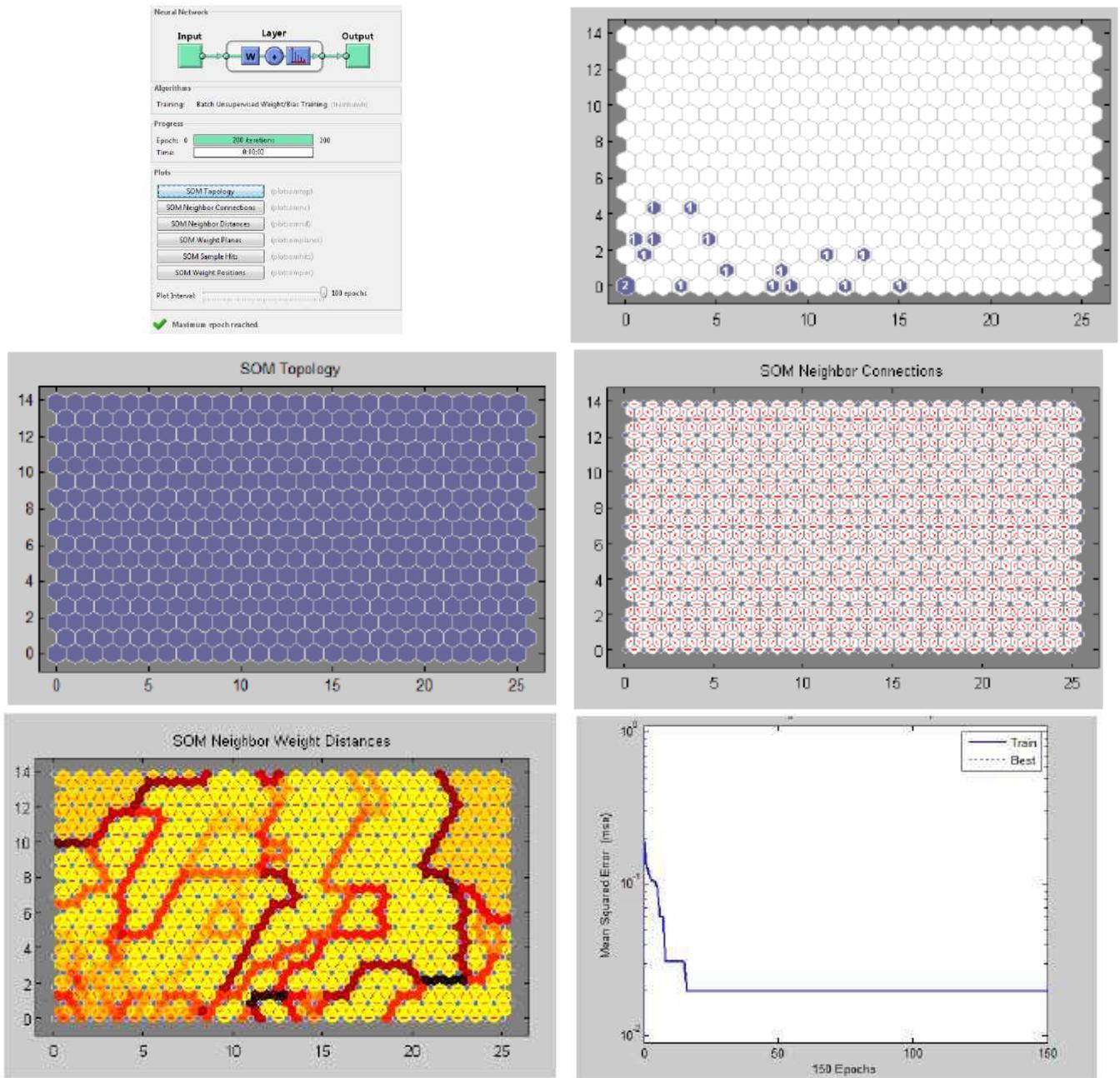


Figura 4. Ejemplo de parametrización de la RNA (arriba izquierda), de la topología de la RNA (arriba derecha), de las conexiones entre neuronas de la RNA (centro izquierda), de las distancias entre neuronas de la RNA (centro derecha), de los planos de pesos de las neuronas de la RNA (abajo izquierda), y de la evolución del error cuadrático medio durante el entrenamiento de la RNA (abajo derecha).

La RNA utilizada para el entrenamiento de los ejemplos obtenidos comenzó con 100 iteraciones, incrementándose de a 50 hasta llegar a 500. En la primera instancia resultados obtenidos representan sólo 2 segmentos correspondientes a las empresas de las ciudades de Rosario y Rafaela, conclusión lograda a partir

del estudio de los diferentes puntos obtenidos y que se aproximan a los núcleos alcanzados. En el resto de las iteraciones se obtuvieron otros segmentos, pero que reflejaban valores puntos que no pertenecían a ningún segmento y que eran núcleos en sí mismos. Se determinó que los datos de la encuesta realizada no reflejaban la situación planteada, contestando a preguntas de control con información cruzada que invalidaba las respuestas. Por ejemplo, en el caso de trabajar con fabricación bajo pedido para luego contestar que la producción realizada era para inventario, reflejando en esta situación, como también en otras preguntas, un antagonismo en las respuestas obtenidas. A partir de la situación detectada, se eliminaron 3 patrones de empresas relevados por considerar que generaban un sesgo en la información obtenida y que, finalmente, en el cálculo del error cuadrático medio aumentaba su valor, impidiendo la generación de agrupaciones representativas a la situación particular de cada empresa.

Como el objetivo propuesto era determinar la segmentación, pero reflejando la relación entre tecnología y procesos industriales en las empresas de cada foco, se continuó con el procedimiento de entrenamiento de la RNA mediante nuevos vectores de pesos creando una RNA con mayor cantidad de neuronas, y especificando el porcentaje de elementos que deberían contener en cada capa de la red y con el vector objetivo y la distancia entre neuronas ya definidos se realizó el entrenamiento, en este caso con 500 iteraciones se obtuvieron 3 segmentos representativos de las empresas de ambas ciudades para las estrategias definidas en la aplicación de tecnologías en sus procesos industriales.

La definición de los umbrales en el procesamiento de datos de separación de cada agrupación se basó en la aproximación mediante pruebas reiteradas de los datos de entrada y su posterior interpretación de los cálculos alcanzados. El tiempo de entrenamiento, utilizando solamente los datos de dos empresas (1 de Rafaela y 1 de Rosario) con seis variables en estudio, demoró menos de 1 segundo, pero en el uso de todos los datos de entrada con todas las variables los tiempos de entrenamiento fueron de varias horas, dependiendo de la cantidad de iteraciones y épocas utilizadas. En el proceso de entrenamiento se realizaron diversos cambios en la topología de la red así como el agregado de mayor cantidad de neuronas hasta lograr la aproximación al resultado mencionado.

5 CLASIFICACION E INTERPRETACION

Los resultados obtenidos a partir de la segmentación lograda se han interpretado a partir de determinar cuáles son las empresas que pertenecen a cada agrupación. Realizando un estudio en detalle de esta información se detectó que la primera comunidad representa a las empresas que incorporaron tecnología, la utilizan en forma masiva para sus procesos industriales y obtienen resultados adecuados en la correlación de tecnología-procesos. En la segunda agrupación, se estableció que las empresas pertenecientes a este segmento representan aquellas que incorporaron tecnología pero su uso es mínimo o no presenta un impacto significativo en el proceso industrial realizado. Finalmente, el tercer grupo constituye el grupo de empresas que no incorporaron tecnologías y cuyo proceso industrial está orientado al uso de herramientas manuales o sin automatizar.

6 CONCLUSIONES

El creciente número de variables presentes en los diversos procesos de fabricación de la industria metalúrgica junto con la complejidad de los procedimientos de producción hace que sea necesario desarrollar herramientas que permitan explotar la gran cantidad de datos para proporcionar información valiosa sobre los cambios en el comportamiento del proceso hacia su mantenimiento u optimización. En este trabajo se propone una nueva visualización de los datos basada en mapas auto-organizativos (SOM) obtenidos a partir de los datos del relevamiento realizado permite explorar diferencias entre dos o más empresas mediante la comparación de la distribución espacial de las neuronas en sus modelos del mapa.

Mediante los resultados obtenidos en esta primera instancia, se muestran indicios de que el estudio y comparación mediante el uso de una RNA es posible en las empresas de nuestra región. La posibilidad de incorporar nuevas empresas y de diferentes sectores productivos aportaría aún más información al entrenamiento de la RNA, permitiendo incrementar y mejorar su segmentación. A partir de la interacción con las empresas relevadas se confirma que valoran seriamente la importancia de una buena gestión de su tecnología aplicada al proceso industrial y su optimización a través de los recursos productivos, por lo que se continuará con el relevamiento de mayor cantidad de datos y a generar diversas estrategias aplicadas a cada segmento en particular. En este caso en particular, para el primer tipo de agrupación, se puede incrementar el uso de la aplicación de nuevas tecnologías e incorporar nuevo equipamiento industrial, mientras que para empresas tipos del tercer segmento se debería generar un plan de sensibilización para la incorporación, adecuación y aplicación de tecnología a los procesos industriales a través de un programa de adaptación del equipamiento industrial. En función del resultado obtenido, y como señalaba J.M. Keynes, "Es mejor estar aproximadamente correcto, que precisamente equivocado" para obtener efectos aplicables tecnológicamente en las industrias de la región.

REFERENCIAS

- Alhoniemi, E., Hollmén, J., Simula, O., & Vesanto, J. (1999). Process monitoring and modeling using the self-organizing map. *Integrated Computer-Aided Engineering*, 6, 3–14.
- Alves, R.M.B., and Nascimento, C.A.O., Analysis and detection of outliers and systematic errors in industrial plant data. *Chemical Engineering Communications*, 194:382-397, DOI:10.1080/00986440600899955, 2007.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection – a survey. *ACM Computing Surveys*, 41, 15:1–15:44.
- Díaz, I., Domínguez, M., Cuadrado, A. A., & Fuertes, J. J. (2008). A new approach to exploratory analysis of system dynamics using SOM applications to industrial processes. *Expert Systems with Applications*, 34, 2953–2965.
- Díaz, I., & Hollmén, J. (2002). Residual generation and visualization for understanding novel process conditions. *Proceedings of international joint conference on neural networks (IJCNN 2002)* (Vol. 3, pp. 2070–2075). Piscataway: IEEE.
- Fornari, J., Grieco, S., Parodi M., y Luccini E., Análisis de líneas de producción utilizando técnicas tradicionales. *Energieia*, 8:41-52, 2010.
- Forst, W., y Hoffmann, D., *Optimization-Theory and Practice*. DOI: 10.1007/978-0-387-78977-4_8. Springer, 2010.
- Gertler, J. J. (1998). *Fault detection and diagnosis in engineering systems* (1st ed.). Marcel Dekker Inc..
- Govindhasamy, J.J., McLoone, S.F., Irwin, G.W., French, J.J., and Doyle, R.P., Neural modelling, control and optimisation of an industrial grinding process. *Control Engineering Practice*, 13:1243-1258, 2005.
- Himberg, J., Ahola, J., Alhoniemi, E., Vesanto, J., & Simula, O. (2001). The selforganizing map as a tool in knowledge engineering. In *Pattern recognition in soft computing paradigm* (pp. 38–65). River Edge, NJ, USA: World Scientific Publishing Co., Inc..
- Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8, 1–8.
- Kohonen, T. (1990). The self-organizing map. In *Proceedings of the IEEE* (pp. 1464–1480). Vol. 78.
- Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Secaucus, NJ, USA: Springer-Verlag New York, Inc.
- Kohonen, T., Oja, E., Simula, O., Visa, A., & Kangas, J. (1996). Engineering applications of the self-organizing map. *Proceedings of the IEEE*, 84, 1358–1384.
- Markou, M., & Singh, S. (2003). Novelty detection: A review-part 1: Statistical approaches. *Signal Processing*, 83, 2481–2497.
- Nomikos, P., & MacGregor, J. F. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics*, 37, 41–59.
- Rao, S.S., *Engineering Optimization: Theory and Practice*, Fourth Edition, John Wiley & Sons, 2009.
- Tryba, V., Metzen, S., & Goser, K. (1989). Designing basic integrated circuits by selforganizing feature maps. In *international workshop on neural networks and their applications Neuro-Nîmes '89* (pp. 225–235). ARC; SEE Nanterre, France: EC2.
- Ultsch, A., & Siemon, H. P. (1990). Kohonen's self organizing feature maps for exploratory data analysis. In *INNC Paris 90* (pp. 305–308). Universitat Dortmund.
- Ypma, A., & Duin, R. P. (1997). Novelty detection using self-organizing maps. In N. Kasabov, R. Kozma, K. Ko, R. O'Shea, Coghill, & T. Gedeon (Eds.). *Proceedings of international conference on neural information processing* (Vol. 2, pp. 1322–1325). Springer.
- Zarenezhad, B., y Aminian, A., Application of the neural network-based model predictive controllers in nonlinear industrial systems. Case study. *Journal of the University of Chemical Technology and Metallurgy*, 46:67-74, 2011.