

IDENTIFICACIÓN DE VARIABLES EN EMPRESAS INDUSTRIALES PARA LA MODELIZACIÓN Y EXTRACCIÓN DE INFORMACIÓN MEDIANTE EL USO DE REDES NEURONALES

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

^aFacultad de Química e Ingeniería “Fray Roger Bacon”, Pontificia, Universidad Católica Argentina,
Av. Pellegrini 3314, 2000 Rosario, Argentina, javier.fornari@uca.edu.ar,
vidaliesteban@hotmail.com, sebastian_grieco@uca.edu.ar, <http://www.uca.edu.ar/bacon>

^bGrupo de Energía Solar, Instituto de Física de Rosario (CONICET-UNRosario), Rosario, Argentina,
luccini@ifir-conicet.gov.ar, parodi@ifir-conicet.gov.ar, <http://www.ifir-conicet.gov.ar>

Keywords: Procesos Industriales, Redes Neuronales, Clasificación, Estrategia Empresarial.

Abstract. Las plantas de los actuales procesos de fabricación industrial están bajo una tremenda presión para mantener y mejorar la calidad del producto y el beneficio en estrictas limitaciones medioambientales y de seguridad. Para un funcionamiento eficiente, cualquier acción de toma de decisiones relacionadas con la operación de la planta requiere el conocimiento del estado real del proceso. La disponibilidad de información y el conocimiento intuitivo de los diferentes estados del proceso tiene un valor incalculable, con implicaciones inmediatas para la rentabilidad, planificación de la gestión, la responsabilidad ambiental y su seguridad. Los procesos industriales exhiben una multitud de estados operativos; algunos de ellos son bien conocidos para los operadores de la planta, mientras que algunos otros no son fácilmente detectables. La detección, análisis y visualización de los diferentes estados de un proceso es esencial y debido a los avances en la tecnología de medición e información, la industria de procesos también está pasando por la era de la explotación de datos. En el marco de estudio de los procesos industriales se identificaron y clasificaron las variables que caracterizan la modelización de las diversas actividades como estrategia para optimizar el funcionamiento de la empresa. Este estudio se aplica en el ámbito de la industria metalúrgica, 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 en este trabajo hacer uso de los datos obtenidos mediante la aplicación de redes neuronales para clasificar y obtener patrones de los diversos procesos. En este caso se hace hincapié en la representación de los resultados obtenidos mediante redes de mapas auto-organizados (SOM, Self Organizing Map). Los métodos que se utilizan están basados en la visualización de las mediciones del proceso y la extracción de información relevante mediante la explotación de la estructura topológica de las observaciones. El objetivo práctico es la identificación de los diversos modos de funcionamiento de la industria y las variables de los procesos más sensibles antes de desarrollar una estrategia de control alternativa. Los resultados preliminares confirman que el enfoque utilizado es capaz de proporcionar valiosa información y ofrece posibilidades para la aplicación directa sobre la industria.

1 INTRODUCCIÓN

Una mejor comprensión y conocimiento acerca de la naturaleza y rendimiento de los procesos industriales es cada vez más necesario para cumplir con los estrictos requisitos de la productividad y la calidad de la industria actual. La mayoría de los procesos industriales incluyen ingredientes como gran número de variables, no linealidades, degradación debido al desgaste o envejecimiento, etc. Estas características aumentan la complejidad de los modelos que se requieren para ser lo suficientemente precisos para sugerir mejoras o para detectar cambios significativos que conduzcan a un impacto en calidad o en la productividad. Por lo tanto, los enfoques basados en modelos pueden no ser de fácil implementación en procesos industriales complejos de pequeñas y medianas empresas industriales. Para este tipo de procesos, por lo general, se carece de modelos confiables o precisos y el conocimiento acerca de la naturaleza de los fallos se expresa rara vez en formas racionales, tal como los modelos matemáticos, o a través de un conjunto bien definido de normas o procedimientos para realizar un diagnóstico.

Los procesos en lote (Nomikos & MacGregor, 1995), caracterizados por su naturaleza finita y estado no estacionario, son de un tipo de proceso que posee fuertes exigencias en materia de supervisión. El objetivo, en este caso, es garantizar que la fabricación del producto especificado sea consistente entre diferentes planes de producción. El enfoque habitual para hacer frente a este problema es el control estadístico de procesos multivariable, que modela el proceso utilizando los datos obtenidos cuando el proceso está en la condición normal de operación. Este enfoque se puede enmarcar como una detección novedosa (Chandola et al., 2009). Los enfoques novedosos de detección de cambios son muy adecuados para este tipo de problemas, donde el conjunto de fallos potenciales es desconocido o incluso inmanejable. Se basan en el cálculo y la evaluación de las diferencias de comportamiento del proceso actual respecto al comportamiento esperado (normal). Por lo general se utilizan las estadísticas (Markou & Singh, 2003) para calcular los umbrales que revelan los cambios significativos en una o más variables del proceso, lo que podría facilitar la detección temprana de fallas. El conjunto de variables que intervienen puede también ser analizado o clasificado (Gertler, 1998) luego usando el conocimiento previo acerca de las fallas. De hecho, los enfoques típicos de clasificación son difíciles de concebir en la supervisión de procesos, ya que requieren datos de entrenamiento bajo todas las condiciones de fallas posibles, las que frecuentemente no están disponibles.

Debido a los avances en los procesos de medición y la tecnología de la información, los procesos industriales están pasando por la era de la explotación de datos ya que los datos históricos están disponibles en abundancia. Los datos derivan a un enfoque basado en el monitoreo de procesos y la supervisión busca construir una representación del proceso a partir de un conjunto de mediciones de fácil acceso sin asumir explícitamente cualquier conocimiento a priori del fenómeno en estudio. Sin embargo, la tarea de conseguir los datos es enorme y sigue siendo una preocupación importante. Es necesario entonces, la disponibilidad de métodos efectivos para modelar los datos para extraer la estructura existente, identificar y reconstruir automáticamente las estructuras más relevantes y permitir una fácil interpretación de las muestras de los diversos estados de la información.

Por otro lado, puede estar disponible una gran cantidad de conocimiento impreciso o no racional, construido sobre la experiencia del personal que invirtió un gran número de horas de observación y supervisión de procesos. Las técnicas de visualización de datos e información (Keim, 2002) son frecuentemente enfoques muy eficientes en tales contextos. La visualización permite utilizar sutiles formas de conocimiento así como tomar la avanzada

habilidad humana para identificar patrones complejos en una forma rápida y sin disponer de recursos de funciones relacionadas a los procesos de razonamiento. Esto hace que el enfoque visual sea muy adecuado para ocuparse del conocimiento no estructurado en complejos procesos industriales.

Los mapas auto-organizativos (SOM) (Kohonen, 2001) han sido exitosamente aplicados como herramienta de visualización de datos en diversos campos (Kohonen et al., 1996) incluyendo la supervisión de procesos (Alhoniemi et al., 1999). También han demostrado ser una técnica poderosa para la detección de patrones (Ypma & Duin, 1997).

2 MAPAS AUTO-ORGANIZATIVOS PARA VISUALIZACIÓN

El mapa auto-organizativo (SOM) (Kohonen, 1990, 2001) es una red neuronal que permite producir un mapeado suavizado no lineal y ordenado de un espacio de entrada multidimensional X en un mapa reticular visual V de menores dimensiones, típicamente en dos dimensiones para su visualización. El espacio de menores dimensiones captura la estructura latente de los datos del espacio de entrada, preservando la topología anteriormente definida en grilla. SOM se basa en un algoritmo de aprendizaje no supervisado y competitivo. Cada neurona o unidad i de la red SOM está representado por las coordenadas g_i de los nodos de una retícula regular en dos dimensiones en V y por un vector de pesos m_i en X . Las neuronas están conectadas en forma adyacente de acuerdo con una relación de vecindad que establece el mapa topológico, generalmente rectangular o hexagonal. Durante el entrenamiento, la red SOM se pliega con el fin de adaptarse a los conjuntos de datos de entrada, y las coordenadas n-dimensionales m_i de las neuronas se actualizan de modo que un mayor número de neuronas se concentran en las zonas con una mayor densidad de datos. Las neuronas dividen el espacio en un conjunto finito de regiones.

El algoritmo original de entrenamiento de una red SOM (Kohonen, 2001) calcula, en cada paso, todas las distancias entre el vector de entrenamiento de entrada x y el vector de pesos m_i . La unidad con la distancia más corta al vector de entrada se denomina unidad de coincidencia más adecuada (Best Matching Unit - BMU) y se denota como c . Los pesos de la unidad m_c , junto con los pesos de sus vecinos son actualizados para aproximar a los datos de entrenamiento, siguiendo un esquema competitivo-cooperativo,

$$c = \arg \min_i \{ \|x - m_i\|^2 \} \quad (1)$$

$$m_i(t+1) = m_i(t) + \alpha(t) h_{ci}(t) [x(t) - m_i(t)] \quad (2)$$

donde $\|\cdot\|$ es la distancia, por lo general Euclidiana, $x(t)$ es el vector de entrada de los conjuntos de entrenamiento en el tiempo t , $h_{ci}(t)$ es la función de vecindad alrededor de la unidad ganadora c y $\alpha(t)$ es el coeficiente de adaptación al tiempo t . La función de vecindad depende del tiempo y la distancia de la unidad i a la neurona ganadora c . El área de influencia de la muestra de entrada en la red SOM se define como:

$$h_{ci}(t) = \exp\left(-\frac{d(g_c, g_i)}{2\sigma(t)^2}\right) \quad (3)$$

donde g_c, g_i son las posiciones de los nodos de las unidades c e i en la grilla SOM, $d(\cdot, \cdot)$ es una métrica de distancia definida en V y $\sigma(t)$ es la amplitud del núcleo de la vecindad. El entrenamiento, usualmente, se realiza en dos etapas. En la primera etapa, los grandes valores

iniciales de los coeficientes del factor de aprendizaje $\alpha(t)$ y la amplitud de la vecindad del núcleo $\sigma(t)$ se utilizan para favorecer un ordenamiento adecuado. En la segunda etapa, tanto el factor de aprendizaje como la amplitud de la vecindad del núcleo se reducen para favorecer a la aproximación de los datos.

En cada etapa de entrenamiento, se obtienen los índices de todas las unidades ganadoras para cada ejemplo x_j y los vectores de pesos cuya partición del conjunto de datos se calcula como:

$$c_j = \arg \min_i \{ \|x_j - m_i\| \} \quad (4)$$

$$m_i(t+1) = \frac{\sum_{j=1}^n h_{c_j i}(t) x_j}{\sum_{j=1}^n h_{c_j i}(t)} \quad (5)$$

donde c_j es el índice de la BMU del ejemplo x_j .

El entrenamiento se realiza para el conjunto de datos y el mapa óptimo se obtiene sobre la base del criterio general de parada, utilizando el error cuadrático medio o la cantidad de iteraciones máxima permitida en el algoritmo.

El algoritmo SOM proporciona maneras poderosas de visualización, ya que crea un mapa suavizado no lineal multidimensional a un espacio de visualización de menores dimensiones (Himberg et al., 2001). Cualquier propiedad escalar significativa del proceso p_i que puede ser evaluada para cada unidad i en el espacio de entrada se puede representar en el espacio de visualización utilizando un píxel centrado en la posición g_i y que tiene un valor de color correspondiente a p_i en una escala de color predeterminado. Para grillas regulares, se producen los planos SOM. Cada plano SOM representa la distribución de una propiedad significativa del proceso para todos los estados del proceso por medio de una imagen en color.

Además, dado que todos los planos se obtienen de la misma manera, todas las imágenes presentan información de una manera consistente, es decir, los planos pueden ser comparados y explorados simultáneamente permitiendo fusionar diferentes tipos de conocimiento sobre el proceso, que van desde simplemente planos que representan las variables de proceso o medidas para los diferentes estados (Tryba et al., 1989) a tipos más elaborados de conocimiento, tales como la información de agrupación (Ultsch & Siemon, 1990), el modelo basado en el conocimiento (Díaz et al., 2004), mapas de correlación de información local (Díaz & Hollmén, 2002) y más recientemente, mapas que muestran las respuestas dinámicas (Díaz et al., 2010).

3 RELEVAMIENTO DE DATOS DE ENTRADA

Para el estudio realizado en este trabajo, se han relevado 43 empresas de la industria metalúrgica, 21 empresas de la zona de Rosario y 22 de la zona de Rafaela, provincia de Santa Fe. En esta sección se reflejará la información referente a cada una de estas empresas ya que en este proyecto en particular se generó un diagnóstico sobre el uso de las Tecnologías de la Información y la comunicación (TICs) en las empresas del sector metalúrgico, con el fin de identificar prioridades y definir inversiones en el área de TICs por parte de las empresas tanto a nivel individual como grupal para confeccionar las estrategias necesarias en las mejoras de sus procesos industriales de fabricación. El objetivo del cuestionario realizado es permitir

conocer las características gerenciales de una empresa indagando en dos sentidos, a) la complejidad tecnológica incorporada a su accionar y al mismo tiempo que se determina b) su situación a nivel productivo.

3.1 Especificación de la encuesta

La encuesta realizada se dividió en siete secciones, con un total de 66 preguntas, que tratan de cubrir las diversas áreas de la empresa para determinar el impacto de las tecnologías en los procesos productivos. Las secciones utilizadas son:

- a. Información de la empresa: datos generales para poder identificar cuál es la empresa a la cual se hace referencia en el estudio a través de la razón social, persona de contacto y producto o servicio principal, entre otros datos (19 preguntas).
- b. Gerencia y Administración: en esta sección los datos solicitados se refieren a la tercerización de diversas áreas o servicios de la empresa así como el nivel de exportación y el uso de indicadores de rendimiento (7 preguntas).
- c. Operaciones: los datos relevados en este apartado se refieren al tipo de insumo que utiliza en su proceso industrial, así como la estrategia de fabricación utilizada (diseño, fabricación y ensamblado bajo pedido o para inventario) y el tipo y frecuencia de mantenimiento del equipamiento industrial utilizado (16 preguntas).
- d. Diseño e Ingeniería de Procesos: en esta sección se solicita la oferta de productos o servicios junto con las mejoras realizadas en los últimos 3 años y el 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 preguntas).
- e. Perfil Innovador: la detección del tipo de perfil de la empresa, así como del empresario y su área ejecutiva, es relevada en esta sección a través de un conjunto de preguntas relacionadas a la visión innovadora de futuro de la empresa y el mercado, fuentes de información relacionadas a la incorporación de tecnologías de la información y comunicación, nivel de educación del personal y capacitaciones realizada en el último año (6 preguntas).
- f. Nivel de adopción de TICs: en este caso las preguntas de este apartado estuvieron dirigidas a la 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 preguntas).
- g. Nivel de utilización de TICs: finalmente, en esta última sección, se determinaron las 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 preguntas).

4 APLICACIÓN AL ANÁLISIS DE DATOS INDUSTRIALES

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 (**¡Error! No se encuentra el origen de la referencia.**) 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.

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	Mono	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 1: Matriz parcial de datos relevados

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 razones, 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 redes neuronales artificiales, en este caso el producto Matlab® versión 7.7.0.471 año 2008.

4.1 Arquitectura utilizada

La red neuronal artificial utilizada para el entrenamiento (**¡Error! No se encuentra el origen de la referencia.**) 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 (**¡Error! No se encuentra el origen de la referencia.**) 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. Del análisis de las empresas que referenciaban se determinó que los datos de la encuesta realizada no reflejaba 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 (**¡Error! No se encuentra el origen de la referencia.**), aumentaba su valor, impidiendo que la generación de agrupaciones representativas a la situación particular de cada empresa.

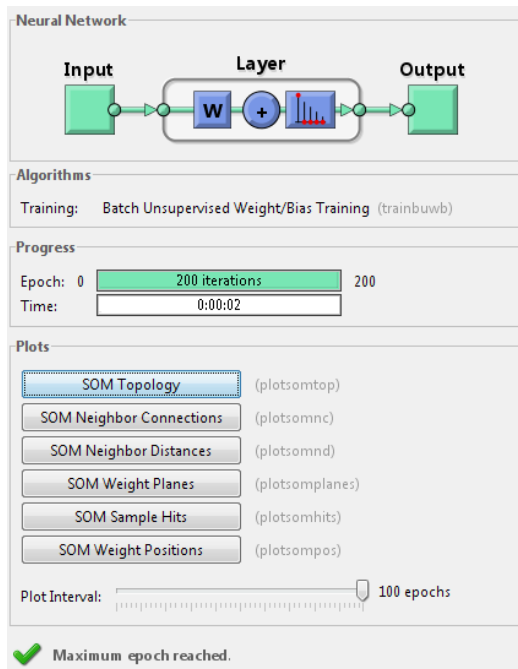


Figura 2: Ejemplo de parametrización de la red

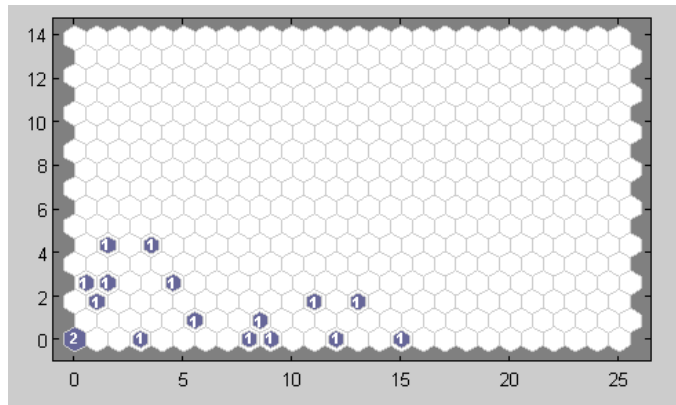


Figura 3: Topología de la red neuronal

Como el objetivo propuesto era determinar la segmentación pero que refleje la relación entre tecnología y procesos industriales en las empresas de cada foco, se continuó con el procedimiento de entrenamiento de la red mediante nuevos vectores de pesos creando una red neuronal con mayor cantidad de neuronas (**¡Error! No se encuentra el origen de la referencia.**), 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 (**¡Error! No se encuentra el origen de la referencia.**) ya definidos se realizó el entrenamiento, y en este caso con 500 iteraciones se obtuvieron resultados que están representados en la **¡Error! No se encuentra el origen de la referencia.** donde se obtuvieron 3 segmentos representativos de las empresas de ambas ciudades para las estrategias definidas en la aplicación de tecnologías en sus procesos industriales.

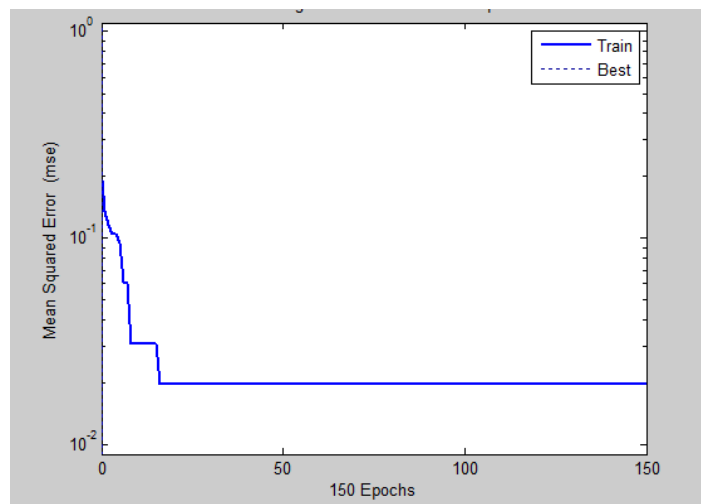


Figura 4: Error Cuadrático Medio

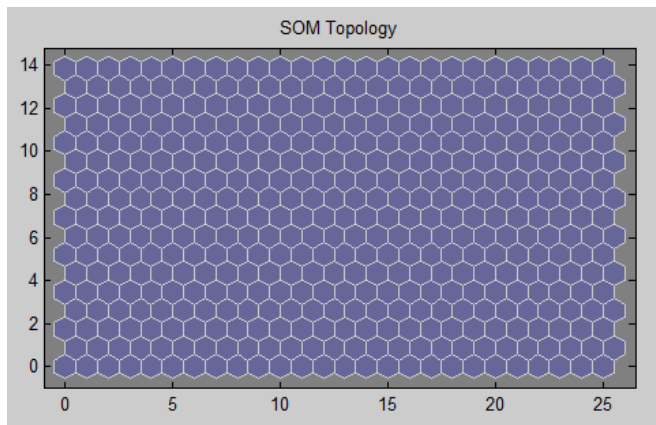


Figura 5: Conexiones entre neuronas

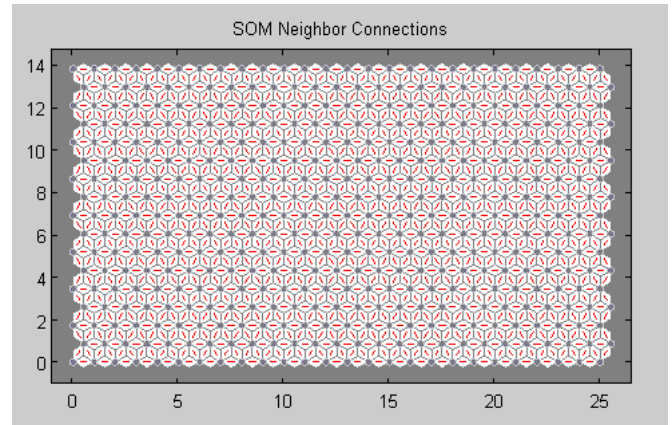


Figura 6: Distancias entre neuronas

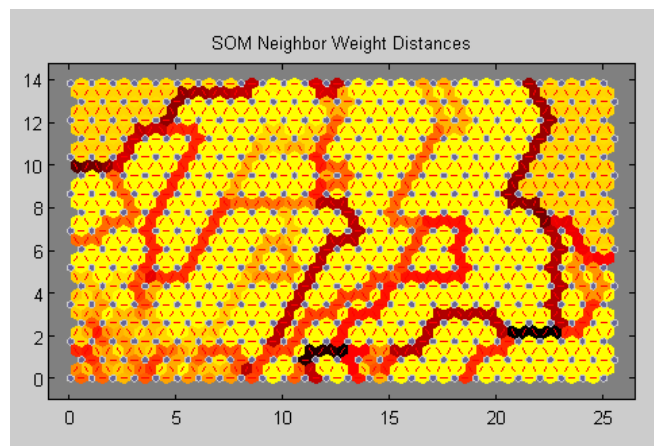


Figura 7: Planos de pesos de neuronas

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 neuronal así como el agregado de mayor cantidad de neuronas hasta lograr la aproximación al resultado mencionado.

4.2 Clasificación e Interpretación

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.

5 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 de empresas mediante el uso de una Red Neuronal Artificial (RNA) es posible en 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.
- 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.
- Gertler, J. J. (1998). *Fault detection and diagnosis in engineering systems* (1st ed.). Marcel Dekker Inc..
- 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.
- 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.