Economía, XXXIII, 26 (julio-diciembre, 2008), pp. 179-188

Una metodología híbrida para el modelo de riesgo proporcional de Cox*

A hybrid methodology for the Cox proportional hazard model

Marianela Luzardo Briceño**

Recibido: 29-02-08 / Revisado: 25-03-08 / Aceptado: 25-04-08

Códigos JEL: F51, L65, N56, N76

Resumen

Se propone una metodología que utiliza la sinergia entre la estadística y la inteligencia artificial para obtener estimaciones de los parámetros del modelo de riesgo proporcional de Cox usando la estructura de la neurona neo difusa propuesta por Yamakawa en 1994. La metodología consta de tres etapas divididas a su vez en fases para obtener las estimaciones en cuestión. Se usaron datos de tres complejos de la industria venezolana CVG-Venalum. La metodología propuesta da resultados con intervalos de confianza más precisos que los obtenidos por la metodología de Cox.

Palabras clave: Modelo, Cox, neurona neo difusa,

Abstract

The paper proposes a methodology that uses the synergy between the statistics and the artificial intelligence to obtain estimations of the parameters of the proportional hazard Cox's Model using the neo diffuse neuron structure proposed by Yamakawa in 1994. The methodology consists of three stages divided in turn in phases to obtain the estimations. There was used information of three complexes of the Venezuelan industry CVG-Venalum. The proposed methodology gives results with confidence intervals more precise that the obtained by the Cox's methodology.

Key words: Model, Cox, neo diffuse neuron

* Esta investigación contó con el apoyo institucional y el financiamiento del Consejo de Desarrollo Científico, Humanístico y Tecnológico (CDCHT) de la ULA en el marco del proyecto E-278-07-02B

** Instituto de Estadística Aplicada y Computación, Universidad de Los Andes, Núcleo La Liria. Edificio G, Mérida, Venezuela. Correo electrónico: nela@ula.ve.

1. Introducción

Esta investigación propone una metodología que utiliza conjuntamente la teoría estadística y los constructos teóricos de la neurona neo difusa (Yamakawa, 1994) para obtener estimaciones de los parámetros del modelo de riesgo proporcional de Cox. Esta metodología híbrida resulta provechosa ya que permite la discretización de las variables en términos de funciones de pertenencia propuestas en la lógica difusa, así la membrecía de un elemento será valorada en función de un rango, y no en términos absolutos, como se propone en las metodologías que se utilizan frecuentemente en la ciencia estadística.

Se justifica en esta investigación la utilización de la neurona señalada en el modelo de riesgo proporcional de Cox porque al realizarse la estimación de Cox, como originalmente se establece, sólo se estaría estimando un modelo lineal que no tomaría en cuenta los posibles ruidos que causan las no linealidades que puedan darse en el proceso productivo, la retroalimentación o la influencia que otras variables del proceso puedan tener sobre una variable en particular. Otra razón por la cual se plantea el uso de la neurona neo difusa es que no es necesario que se cumpla el supuesto de proporcionalidad (el riesgo entre los componentes permanece constante a lo largo del tiempo) en los datos para aplicarlas.

Varios autores (Faraggi y Simon, 1995; Ravdin y Clark, 1992; Ohno-Machado *et al*., 1995) han incursionado en el mundo de la inteligencia artificial para estimar los parámetros del modelo de riesgo proporcional de Cox a través de las redes neuronales artificiales y han obtenido resultados satisfactorios. Sin embargo, al utilizar esta técnica se tiene la desventaja de no saber a ciencia cierta, ni el número de capas ocultas ni el de neuronas por capa que tendría la red neuronal, lo que origina así el uso excesivo del tiempo de cómputo. Otra debilidad de las redes neuronales artificiales se presenta durante el entrenamiento ya que es posible conseguir un error que sea mínimo, para un espacio de tiempo, lo que se consideraría un error local. Sin embargo, este error puede no ser el global. Estos inconvenientes serán solventados a partir de la neurona neo difusa.

Ahora bien, este estudio propone que en el modelo de riesgo proporcional de Cox sea incorporada la salida de una neurona neo difusa, es decir, que sea sustituida la concepción lineal que subyace en este modelo por una noción no lineal. De esta forma, se estaría dando cabida al postulado de la lógica difusa de que "todo es cuestión de grado" (Zadeh, 1965), así los elementos a estudiar estarían en varios conjuntos de acuerdo con su grado de membrecía. Este grado permitirá al investigador manejar fenómenos naturales que son imprecisos, es decir, que tienen implícito un cierto grado de difusidad en la descripción de su naturaleza.

2. Los datos

La empresa Industria Venezolana de Aluminio, C.A. (CVG-Venalum) fue creada el 29 de agosto de 1973 con el objeto de producir aluminio primario con fines de exportación. Está ubicada sobre la margen del río Orinoco, en la ciudad de Puerto Ordaz, Estado Bolívar, al sur de Venezuela, y constituye la mayor planta reductora de aluminio primario en Latinoamérica con una capacidad instalada de cuatrocientas treinta mil toneladas por año.

CVG-Venalum cuenta con cinco líneas de producción de aluminio, cuatro que utilizan tecnología Reynolds P-19 y una quinta que usa tecnología Hydro-Aluminium. El aluminio producido viene presentado en lingotes, cilindros para extrusión y aluminio líquido. Las celdas que utilizan tecnología Reynolds P-19, y sobre la cual se basa el estudio, se identifican porque el sistema de alimentación de alúmina está compuesto por cuatro alimentadores con su respectivo rompecostra que operan independientemente. Cada celda usa 18 ánodos con una vida útil de 22 días cada uno de ellos y una capacidad útil de producción mensual de 36 toneladas de aluminio por celda. La temperatura de operación de la celda es 960 ^{o}C, la adición de fluoruro de aluminio es manual y el voltaje de operaciones 162 KA. La frecuencia de trasegado es cada 24 horas y la subida de puente es realizada cada 15 días [5].

Se tomó una muestra de trescientas (300) celdas electrolíticas pertenecientes a los complejos I, II y III de la empresa venezolana pro-

ductora de aluminio primario CVG-Venalum -Puerto Ordaz, estado Bolívar, con un error de estimación del 5,7%. Para lograr una mejor representación de la población, el tamaño muestral se distribuyó equitativamente para cada complejo, dentro de los cuales se tomó una muestra aleatoria simple (con reposición) de tamaño cien. El evento de interés que se tomó en cuenta fue falla por perforación en el cátodo y el periodo de estudio abarca desde enero de 1998 hasta septiembre de 2004. De las trescientas celdas que conformaron la muestra, doscientas setenta y nueve presentaron el evento de interés y veintiuna fueron censuradas.

Por otro lado, se generaron 50 muestras aleatorias utilizando el procedimiento estadístico de *bootstrap* (método de remuestreo propuesto por Efron en 1979) donde a cada una de las muestras generadas se procedió a aplicarle la metodología anteriormente descrita para obtener intervalos de confianza para cada uno de los parámetros estimados y poder así compararlos con los obtenidos por el procedimiento propuesto por Cox.

3. Metodología híbrida

En este apartado se presenta la Metodología Hibrida para el Modelo de Riesgo Proporcional de Cox (1972) la cual consta de tres etapas que se encuentran subdivididas a su vez en fases.

3.1 Etapa I: Selección del protocolo experimental

En la etapa I se describe la naturaleza del protocolo experimental –la unidad de estudio, la selección de variables, la definición del evento de interés y la temporalidad, a partir de un proceso cualquiera de producción industrial. En esta etapa se contemplan cuatro fases.

- *Fase 1*. Análisis y descripción del proceso a investigar.
- *Fase 2*. Selección de variables que intervienen en el proceso a investigar.
- *Fase 3*. Definición del evento de interés (falla).
- *Fase 4*. Temporalidad del estudio.

3.2 Etapa II: Construcción de la estructura matemática de la neurona neo difusa

En la etapa II se estiman los pesos de la neurona neo difusa a partir de la función de verosimilitud parcial (Cox, 1972) para la construcción de la función matemática que la exprese.

- *Fase 1*. Definición del número de conjuntos difusos de las variables que intervienen en el proceso a investigar.
- *Fase 2*. Representación matemática de las funciones de membrecía para cada una de las variables de entrada.
- *Fase 3*. Obtención de las sinapsis no lineales

$$f_j(X) = \mu_{j1}(X)w_{x1} + \mu_{j2}(X)w_{x2} \tag{1}$$

Donde $\mu_{j1}(X)$ y $\mu_{j2}(X)$ son los grados de pertenencia a los diferentes conjuntos difusos.

- *Fase 4*. Obtención de la salida de la neurona neo difusa para cada patrón de entrada
- *Fase 5*. Obtención de los pesos de la neurona neo difusa utilizando el método de Máxima Verosimilitud Parcial

3.3 Etapa III: Aplicación del vector de salida de la neurona neo difusa en el exponente del modelo de riesgo proporcional de Cox

En la etapa III, entrenada ya la neurona, debido a que se conocen todos sus valores, es posible entonces dar paso a la adecuación del modelo de riesgo proporcional de Cox. Esta propuesta metodológica parte de la premisa de que el vector de salida puede ser equivalente al exponente del modelo de riesgo proporcional de Cox.

- *Fase 1*. Construcción del sistema de ecuaciones: se plantea sustituir las salidas de los modelos matemáticos de la neurona neo difusa entrenada por máxima verosimilitud en el exponente de la función de

riesgo $h(t/x_i) = h_0(t)e^{x_i'\beta}$ propuesta por Cox (1972) y se obtiene:

$$h(t/\boldsymbol{x_i}) = h_0(t)exp(g(x_i,\theta)) \quad (2)$$

Y se consigue el siguiente sistema de ecuaciones:

$$S_{nx1} = X_{nxp}\,\beta_{px1} \quad (3)$$

Donde ***S*** es el vector que contiene las salidas emanadas por la neurona neo difusa cuyos parámetros fueron obtenidos por máxima verosimilitud.

- *Fase 2.* Obtención de las estimaciones de los parámetros del Modelo de Regresión de Cox.
  Se determina en esta fase un vector β' que, al sustituir en el miembro derecho del sistema de ecuaciones, dado en la Etapa III, Fase 1, genera un nuevo miembro izquierdo, llámese **S'**, que tiene la propiedad de minimizar la diferencia entre de **S'** y **S**. Es decir, la suma de los cuadrados de los errores cometidos $(S_i' - S_i)$ en cada ecuación, es mínima.

4. Resultados

A continuación se presentan los resultados de las estimaciones de los coeficientes del modelo de riesgo proporcional de Cox a través de las dos metodologías. Esta comparación no pretende ser prescriptiva sino más bien descriptiva, debido a que procura evidenciar el comportamiento de las estimaciones por lo métodos empleados en esta investigación, sin en ningún caso emitir juicios de valor.

Se ha considerado necesario comenzar la comparación con una representación gráfica de las estimaciones obtenidas por el modelo de riesgo proporcional de Cox y la metodología híbrida, para de esta forma ser más claros en la exposición. Si se observa la figura 1, no hay que perder de vista que las estimaciones de los parámetros de la mayoría de

las variables no difieren en forma significativa entre ambos métodos, excepto para la variable corriente que resultó ser significativamente mayor en la estimada por el modelo proporcional de Cox.

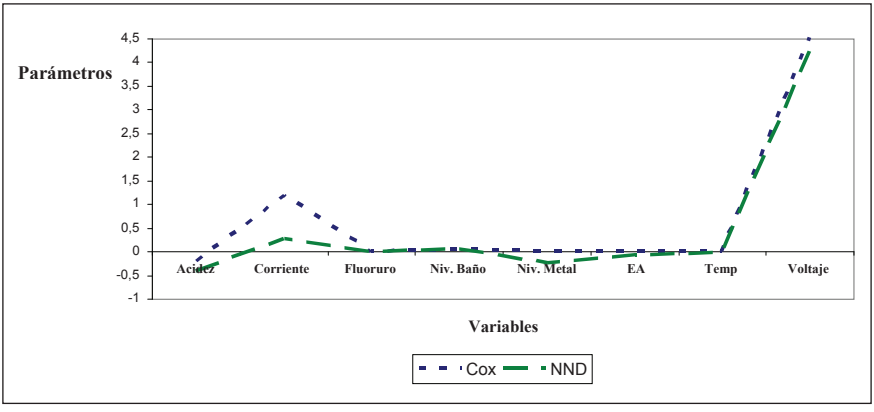

Figura 1. Comparación de las estimaciones entre el Modelo Proporcional de Cox y la metodología

Para continuar con la comparación de las estimaciones a partir de los dos métodos utilizados por esta investigación es necesario analizar y comparar la longitud de los intervalos de confianza de cada uno de ellos para conseguir de este modo la eficiencia relativa de los mismos. Estos valores se presentan en el cuadro 1.

Esta tabla muestra entonces los intervalos de confianza para cada una de las estimaciones obtenidas por el modelo de riesgo proporcional de Cox y por la metodología híbrida así como también la eficiencia relativa entre los dos procedimientos. Es notorio en este cuadro que la metodología híbrida produce intervalos de confianza más estrechos por lo tantos más eficientes y precisos que los obtenidos por método de regresión Cox.

Por otro lado, el cuadro 2 ilustra las estimaciones de los parámetros obtenidos por el modelo de riesgo proporcional de Cox y la metodología híbrida de aquellas variables que no estuvieron correlacionadas y que presentaron como evento de interés *perforación en el cátodo.*

Cuadro 1. Eficiencia Relativa de la Longitud del Intervalo

	Modelo de Cox			Metodología Híbrida			Eficiencia Relativa
	Intervalo de Confianza		Longitud IC	Intervalo de Confianza		Longitud IC	
	LI	LS		LI	LS		
Acidez	-0,2519	-0,1308	0,1211	-0.5186	-0.31925	0,1993	1,65
Corriente	0,9715	1,4176	0,4461	0.1425	0.3871	0,2446	0,55
Fluoruro	-0,0121	0,0032	0,0153	-0.0074	-0.0055	0,0019	0,12
Nivel deBaño	0,0082	0,1129	0,1047	0.0395	0.0533	0,0138	0,13
Nivel de Metal	-0,1826	-0,0827	0,0999	-0.3022	-0.1863	0,1159	1,16
Efectos Anódicos	-0,2168	0,0957	0,3125	-0.0797	-0.0647	0,015	0,05
Temperatura	-0,0186	-0,0028	0,0158	-0.0138	-0.0111	0,0027	0,17
Voltaje	1,7476	7,3209	5,5733	3.5532	4.8484	1,2952	0,23

Cuadro 2. Estimación de los Coeficientes a partir de las Metodologías de Cox e Híbrida

	Metodología de Cox		Metodología Híbrida	
Variables	Coeficiente (β)	Exp(β)	Coeficiente (β)	Exp(β)
Acidez	-0,19000	0,827	-0,423139	0,654988
Corriente	1,19464	3,302	0,259825	1,296703
Fluoruro	-0,00446	0,996	-0,0071462	0,992872
Nivel de Baño	0,06060	1,062	0,0467535	1,047864
Nivel de Metal	-0,13268	0,876	-0,244316	0,783245
Efectos Anódicos	-0,06055	0,941	-0,073637	0,929009
Temperatura	-0,01079	0,989	-0,012353	0,987723
Voltaje	4,53400	93,158	4,236206	69,145017

Fuente: elaboración propia

5. Conclusiones

Se evidencia que la discretización de las variables en términos de funciones de pertenencia permite obtener parámetros de la función de riesgo de una manera más eficiente que el modelo de riesgo proporcional de Cox tomando como referencia para argumentar este hecho la amplitud de los intervalos de confianza.

Sobre la aplicación de la metodología híbrida –al caso específico de la empresa bajo estudio– se concluye que las variables corriente, nivel de baño y voltaje presentaron mayor riesgo de desincorporación de la celda por perforación en el cátodo, por lo tanto, se le sugiere a la empresa dar seguimiento y control sobre estas variables con el fin de minimizar el riesgo de desincorporación de las celdas.

Se evidencia de esta manera la necesidad imperiosa de continuar haciendo investigación que permita crear metodologías con el fin de reconocer no sólo avances en las industrias sino también para que se constituyan en ahorros para las mismas, lo que se constituye en un alto beneficio no sólo para la industria en general sino también para la sociedad en particular. Se propone trabajar con datos simulados bajo condiciones ideales para corroborar los resultados obtenidos en este trabajo.

6. Referencias

Cox, D.R. (1972). Regression Models and Life Tables (with Discussion) *J. R. Statistics Soc*. B 34, pp. 187-220.

Cox, D.R. (1975). "Partial Likelihood." *Biometrika*, 62, pp. 269-276.

Cox, D.R. (1979). "A Note on the Graphical Analysis of Survival Data*." Biometrika*, 66, pp. 248-275.

Cox, D.R. and Oakes, D., (1994). *Analysis of Survival Data,* Chapman & Hall, London

CVGVALUM, (Marzo, 2007). http://www.unet.edu.ve/~apuello/CVGVALUM.doc.

Efron, B. (1979). "Bootstrap methods: Another look at the Jacknife*." Ann Statist*. 7: 1-26.

Yager, R. y Zadeh, L., (1992). *An Introduction to Fuzzy logical Applications in Intelligent Systems,* London: Kluwer Academic Publishers.

Yamakawa, T. (1994). *A Neo Fuzzy Neuron and Its Applications to System Identification and Prediction of Chaotic Behavior*. IEEE, Vol. 3 383-395.

Zadeh, L. A., (1965). "Fuzzy Sets." *Information and Control*, vol. 8, pp. 338-353.