

Modelo de máquinas de vectores de soporte para regresión aplicado a la estimación de la tensión de ruptura por termofluencia en aceros ferríticos

Support vector machine model for regression applied to the estimation of the creep rupture stress in ferritic steels

*Carlos Alberto Donis Díaz**, *Eduardo Valencia Morales*, *Carlos Morell Pérez*

Centro de Estudios de Informática (CEI). Universidad Central de Las Villas.
Carretera Camajuaní, Km 51/2 CP. 54830, Santa Clara, Villa Clara, Cuba

(Recibido el 28 de marzo de 2008. Aceptado el 6 de noviembre de 2008)

Resumen

Teniendo como antecedente el empleo de Redes Neuronales en el pronóstico de la tensión de ruptura por termofluencia (*creep*) en aceros ferríticos, en el presente trabajo se realizan nuevos experimentos, utilizando un método de reciente desarrollo dentro del campo del aprendizaje automatizado: las Máquinas de Vectores de Soporte para Regresión (SVMR). Se realizó un análisis comparativo entre ambos métodos obteniéndose resultados satisfactorios por parte de este último. Los resultados son fundamentados teóricamente proponiéndose al final, el empleo de un modelo de SVMR que utiliza un kernel polinomial de grado 3 y constante de regularización igual a 100 para estimar la tensión de ruptura por *creep*.

----- **Palabras Clave:** Termofluencia, aceros ferríticos, máquina de vectores de soporte, redes neuronales.

Abstract

Having as antecedent the use of artificial neural networks (ANN) in the estimation of the creep rupture stress in ferritic steels, new experiments have been developed using Support Vector Machine for Regression (SVMR), a recently method developed into the machine learning field. A comparative analysis between both methods established that SVMR have a better behavior in the problematic of creep. The results are explained theoretically and finally, the use of a model of SVMR that uses a polynomial kernel of third grade and a control capacity constant of 100, is proposed.

----- **Keywords:** Creep, ferritic steels, support vector machine, artificial neural network.

* Autor de correspondencia: teléfono: + 53 + 42 + 281 51 5, fax: + 53 + 42 + 281 6 08, correo electrónico: cadonis@uclv.edu.cu (C. Donis).

Introducción

Durante más de medio siglo, en muchas aplicaciones que abarcan campos tan disímiles como la industria petroquímica, energética y aeronáutica, se han utilizado extensivamente aceros ferríticos resistentes a la termofluencia (*creep*) como es el caso del Fe-2.25Cr-1Mo y Fe-(9-12)Cr. Lo anterior se debe a la excelente fiabilidad que estos aceros muestran para condiciones muy agresivas y su uso durante largos períodos de servicio (como por ejemplo 30 años). En la historia del diseño de los aceros se puede observar un desarrollo progresivo, particularmente en el sentido de soportar altas temperaturas de vapor al utilizarse en plantas de energía contribuyendo a que éstas operen con gran eficiencia, o por ejemplo, en turbinas de aviones garantizando un gran nivel de seguridad. Esta idea permite concluir que los principios básicos del diseño de aceros resistentes a la termofluencia, están bien establecidos y bien fundamentados en base a la experiencia. Los aceros deben tener una microestructura estable que contenga finos carburos aleados resistentes a los movimientos de las dislocaciones; sin embargo, es inevitable que durante largos períodos de servicio o condiciones muy críticas, se produzcan cambios. Es necesario, por tanto, garantizar un potencial de endurecimiento por solución sólida suficiente que permita mantener de forma adecuada la microestructura y asegurar así una elevada resistencia a las deformaciones por termofluencia a largo plazo. También se necesitan otros requerimientos como es la soldabilidad y resistencia a la corrosión y la oxidación. Teniendo en cuenta estas ideas se puede deducir la gran cantidad de variables que intervienen en el diseño de los aceros para lograr determinadas cualidades y propiedades mecánicas y lo difícil que resulta poder expresar alguna relación cuantitativa o interacción que tiene lugar entre dichas variables durante el tiempo de servicio de estos aceros. A lo largo del desarrollo de los aceros, se han utilizado varios métodos empíricos para estimar la tensión de ruptura por termofluencia destacándose recientemente, el empleo de redes neuronales. En este sentido se han desarrollado varios trabajos

basados en un modelo de red neuronal conocido como Multilayer Perceptron (MLP) el cual ha sido entrenado en un ambiente Bayesiano desarrollado por D. Mackay [1] y ha demostrado ser superior en la extrapolación y representación de datos de *creep* [2]. Estudios recientes [3] muestran resultados de las predicciones realizadas con este modelo. En el presente trabajo, a partir de un estudio comparativo, se propone un nuevo método basado en máquinas de vectores de soporte para regresión (SVMR por sus siglas en inglés de *Support Vector Machine for Regression*), modelo de reciente desarrollo dentro del campo del aprendizaje automatizado para problemas de estimación de funciones continuas. Mediante argumentos teóricos y resultados de los experimentos realizados se exponen las ventajas de este método al compararlo con modelos basados en redes neuronales artificiales (RNAs)

Nuevos experimentos en el pronóstico de la tensión de ruptura por creep

Datos utilizados

Para los experimentos se utilizó una recopilación de casos que se encuentran expuestos en el sitio web MAP [4] (siglas en inglés de *Materials Algorithms Project*) proyecto desarrollado por el Laboratorio Nacional de Física y la Universidad de Cambridge, UK. Consiste de 2066 casos que muestran los valores de la tensión de ruptura por *creep* relacionados con 37 variables y compuestos principalmente por valores de dos tipos de aceros típicos: el Fe-2.25Cr-1Mo y el Fe-(9-12)Cr. Estas variables incluyen información sobre la composición química de los aceros (16 componentes), los tratamientos térmicos así como de los tiempos y temperaturas de servicio. Una caracterización de estos datos se puede observar en [5, 6] donde fueron utilizados anteriormente en experimentaciones sobre esta problemática.

Métodos utilizados

A continuación serán referenciadas consideraciones de interés sobre los métodos utilizados en el

presente trabajo con vistas al análisis de los resultados obtenidos.

Redes Neuronales

En [5, 6] se puede observar una descripción detallada de experimentos realizados para pronosticar la tensión de ruptura por *creep*, empleando redes neuronales desarrolladas en un ambiente bayesiano.

En nuestros experimentos se utilizó un MLP formado por 37 nodos en la capa de entrada correspondiendo con cada una de las variables mencionadas en la sección 2.1, una capa oculta con un número variable de nodos (se experimentó con cantidades desde 10 hasta 60) y una capa de salida con un solo nodo cuya salida corresponde al estimado de la tensión de ruptura por *creep*. La función de transferencia presente en los nodos de la capa oculta, al igual que en los trabajos referenciados, fue la tangente hiperbólica. El método de entrenamiento utilizado fue el *Backpropagation* con 10000 iteraciones como máximo.

Máquinas de vectores de soporte para regresión (SVMR)

Una descripción detallada de este método puede ser revisada en [7]. A los efectos del presente trabajo es imprescindible destacar la función de coste a minimizar en la cual se fundamenta este método. SVMR adiciona al término de control del riesgo empírico en la función de coste (único término empleado en las redes neuronales), un término de control de capacidad que es la norma Euclidiana $\|w\|^2$

De tal forma el problema de optimización que se plantea queda expresado como:

$$\text{minimizar } \tau(w, \xi^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \quad (1)$$

sujeto a:

$$f(x_i) - y_i \leq \varepsilon + \xi_i, \quad y_i - f(x_i) \leq \varepsilon + \xi_i, \quad \xi_i^* \geq 0$$

Donde el término $\sum_{i=1}^m (\xi_i + \xi_i^*)$, se refiere al riesgo empírico, $\frac{1}{2} \|w\|^2$ se refiere al nuevo término de control de capacidad incluido y C es una constante

denominada de regularización que establece un compromiso entre la llanura de la función y la cantidad hasta la cual se toleran desviaciones mayores que ε .

Para la experimentación con el modelo de SVMR utilizado en el presente trabajo se emplearon tres tipos de *kernel* diferentes con las siguientes características:

- Kernel Polinomial: $K(x,y) = \langle x, y \rangle^p$ donde p es el orden del polinomio.
- Kernel Polinomial Normalizado:

$$K(x,y) = \frac{\langle x, y \rangle}{\sqrt{\langle x, x \rangle \langle y, y \rangle}}$$
 donde,
 $\langle x, y \rangle = \text{Kernel Polinomial}(x, y)$
- Kernel de Funciones de Base Radial:

$$K(x,y) = \exp(-\sigma * \langle x - y, x - y \rangle^2)$$

Diseño de los experimentos

De la base de casos original se obtuvieron 30 bases de casos con similar tamaño (2066 casos) mediante un proceso de selección o muestreo aleatorio sin duplicados. De cada base de casos se tomaron los primeros ejemplos (un 80%) para entrenar y obtener el modelo y el resto de los casos (20%) para probar el modelo, de manera tal que durante el proceso de entrenamiento o “aprendizaje”, los casos de prueba nunca fueron “vistos”.

Para cada método utilizado en la experimentación se hicieron 30 corridas correspondientes a cada una de las bases de casos obtenidas.

La aplicación de los métodos se logró a partir de la implementación que tiene de los mismos, el ambiente de trabajo WEKA (Waikato Environment for Knowledge Analysis, plataforma de trabajo que agrupa un gran número de técnicas y algoritmos de inteligencia artificial para el análisis de datos) en su versión 3.5.5. En la ayuda de este sistema se puede encontrar una breve descripción de la implementación de los mismos.

Para cuantificar el comportamiento de los métodos se utilizó una medida del error en la estimación de la tensión de ruptura por *creep* en el con-

junto de datos de prueba comparando los valores estimados de salida y_μ con los valores provenientes de las mediciones experimentales t_μ . Específicamente se empleó la raíz del error cuadrático medio (*rmse* – root mean squared error) que brinda valores en el orden de los datos que se desean estimar. Este indicador se obtiene como:

$$rmse = \sqrt{\frac{\sum_{\mu=1}^m (t_\mu - y_\mu)^2}{m}}, \quad (2)$$

donde m es la cantidad de ejemplos del conjunto de prueba.

En el procesamiento estadístico se utilizó el test de Friedman para obtener un ranking de las mejores configuraciones dentro de cada método y

luego el test de Wilcoxon para realizar una comparación par a par de los resultados obtenidos y determinar si existían diferencias significativas.

Análisis y discusión de resultados

En la Tabla 1 se puede observar la configuración utilizada para cada método y sus resultados reflejándose el promedio de la raíz del error cuadrático medio (*rmse*) obtenido a partir de las 30 corridas. Los métodos aparecen ordenados según el ranking obtenido del test de Friedman y se muestra además el nivel de significación estadística entre los resultados. Se incluye el resultado obtenido en el trabajo de F.Brun [5] (donde se utilizó el mismo conjunto de datos que en el presente trabajo) para tener una idea comparativa con los resultados actuales.

Tabla 1 Resultados de los métodos

Nro	Método – Configuración	RMSE promedio	Ranks según Friedman	Diferencia Significativa según test de Wilcoxon		
				* existe – no existe	I	II
I	SVMR con kernel polinomial grado 3, C = 100	12,47	1,00			
II	SVMR con kernel polinomial grado 2, C = 100	14,95	2,00	*		
III	SVMR con kernel polinomial grado 2, C = 1	15,11	2,53	*	*	
	MLP ambiente bayesiano [5]	16,57 ⁽¹⁾				
IV	MLP con BackPropagation, 11 nodos ocultos, 10000 iter	18,59	3,61	*	*	*

⁽¹⁾ Este valor se obtuvo a partir de las gráficas expuestas en [5], donde el mejor valor obtenido del indicador e ($e = \frac{1}{2} (\sum (t_u - y_u)^2)$) fue de 0,154 lo cual significa un valor de raíz del error cuadrático medio (*rmse*) de 16,57 si consideramos $m = 1033$, utilizamos la forma de normalización descrita en dicho trabajo y el cálculo de *rmse* se realiza según (2)

Al comparar los resultados del modelo de red neuronal de mejores resultados (MLP con 11 neuronas en la capa oculta y entrenado con *BackPropagation*) con el método SVMR se puede observar que este último presenta un mejor comportamiento. Los motivos de este resultado pueden estar relacionados con el dilema varianza – sesgo, presente tanto en estimadores paramé-

tricos como no paramétricos. Para que una red neuronal realice un ajuste óptimo, el número de ejemplos de entrenamiento debería tender a infinito [8], pues para un conjunto pequeño, suele ser muy sensible a casos particulares de pares entrada-salida seleccionados para realizar el aprendizaje. La causa es que la red neuronal, estimador de modelo libre, posee inherentemente una gran

varianza. En la literatura [9] se ha corroborado una forma de inducir la cantidad óptima de ejemplos de entrenamiento en dependencia del orden del error de generalización (ε) con que se desee estimar y la cantidad de parámetros libres a ajustar en la red (w). Se plantea que debe estar en el orden de w/ε , en este caso, para $\varepsilon=0,1$ (un 10%) y la topología de mejor comportamiento (37 nodos de entrada, 11 nodos en la capa oculta y un nodo de salida) que contiene 430 parámetros libres entre pesos y umbrales serían necesario 4300 ($430 * 10$) ejemplos de entrenamiento para realizar un ajuste óptimo. La única forma de controlar la elevada varianza que la red neuronal posee inicialmente es introducir en su arquitectura algún tipo de sesgo o información apriorística sobre el problema a resolver. Este problema puede ser abordado con técnicas de regularización que mediante la introducción del sesgo, provocan que el *mapping* que implementa la red neuronal sea suave, es decir, que a entradas similares haga corresponder resultados próximos. Este sesgo puede ser introducido en la función coste en forma de términos adicionales $\varphi(w)$ que miden la desviación de los resultados actuales respecto de la restricción planteada. Muchos algoritmos como las redes de regularización y las redes neuronales con decadencia de pesos, minimizan una expresión de riesgo similar a la siguiente:

$$R[ff] = R_{emp}[ff] + \frac{\lambda}{2} \|w\|^2 \quad (3)$$

siendo λ el denominado parámetro de regularización, que controla el compromiso entre el grado de suavidad de la solución frente al nivel de ajuste de los datos de entrenamiento que alcanza el modelo. Estas características de regularización, están implícitas en la concepción del modelo de SVMR que se basa en el principio de minimización del riesgo estructural al introducir en la función de coste un término de control de capacidad ($\|w\|^2$), véase la función a minimizar (1) descrita en la sección 2.2.2; esto hace esperar un mejor resultado. Minimizar (1) es equivalente a minimizar (3) si asumimos que $C = 1/(\lambda m)$.

Teniendo en cuenta el análisis anterior podemos fundamentar los resultados obtenidos en los ex-

perimentos. Se observa que el mejor valor de *rmse* obtenido por MLP fue de 18,59 para el caso de una topología con 11 nodos ocultos, este valor resulta significativamente diferente (obsérvese el resultado del análisis estadístico) al obtenido por SVMR para el kernel polinomial de grado 2 y $C = 1$ que fue de 15,11 (evidentemente, un mejor valor). Este resultado permite concluir que puede existir un cierto nivel de sobreajuste en el caso del entrenamiento del MLP, hipótesis probable si se analiza además el hecho de no contar con una cantidad óptima de ejemplos de entrenamiento para alcanzar niveles de errores bajos. Este aspecto no afecta a SVMR que mediante el elemento de control de capacidad, hace que ejemplos con posible ruido, no influyan en la estimación de la función final y por tanto permite obtener mejores resultados. Al aumentar el valor de C desde 1 hasta 100 estamos permitiendo que la función a estimar sea menos llana y exista una tolerancia menor a los errores mayores que ε (en los experimentos se consideró 0,01). Como se puede observar se logra una mejoría en el valor de *rmse* = 14,95 lo cual reafirma el mejor comportamiento de este método al compararlo con el MLP descrito para la problemática del *creep* realizando aproximaciones mas exactas sin llegar a permitir sobreajuste.

Un elemento importante en la utilización de SVMR es el *kernel* seleccionado. En nuestros experimentos se obtuvo el mejor resultado con el *kernel* polinomial con grado 3 cuyo *rmse* fue de 12,47 obteniéndose una diferencia significativa según el test de Wilcoxon incluso al compararlo con el resultado del *kernel* polinomial de grado 2, segundo modelo de mejores resultados.

Conclusiones

El empleo de SVMR en el pronóstico de la tensión de ruptura por *creep* resulta un método más eficiente que el empleo de las redes neuronales ya que permite realizar estimaciones con un nivel de error menor. En lo anterior influye el hecho de que en la construcción del modelo de pronóstico, se elimina la posibilidad del sobreajuste por la influencia de datos con posible ruido o no carac-

terísticos. Un modelo que utilice un *kernel* polinomial de grado 3 y control de capacidad $C=100$ permite obtener los mejores resultados.

Referencias

1. D. J. C. MacKay. "Bayesian Methods for Neural Networks: Theory and Applications". *Neural Networks Summer School. Cambridge University. U.K.* 1995. pp. 15-24.
2. H. K. D. H. Bhadeshia, T. Sourmail. *Japan Society for the Promotion of Science, Committee on Heat-Resisting Materials and Alloys*. Vol. 44. 2003. pp. 299-314.
3. F. Masuyama, H. K. Bhadeshia. "Creep strength of high CR Ferritic Steels designed using neural networks and phase stability calculations". *Fifth International Conference on Advances in Materials Technology for Fossil Power Plants October 3-5 (2007)*. 4B-01. EPRI. Palo Alto. California. 2007.
4. <http://www.msm.ac.uk./map/map.html> Consultada el 20 de enero de 2008.
5. F. Brun, T. Yoshida, J. D. Robson, V. Narayan, H. K. D. H. Bhadeshia, D. J. C. MacKay. "Theoretical design of ferritic creep resistant steels using neural network, kinetic, and thermodynamic models". *Materials Science and Technology*. Vol. 15. 1999. pp. 547-554.
6. D. Cole, C. Martin-Moran, A.G. Sheard, H. K. Bhadeshia, D. J. C. MacKay. "Modelling creep rupture strength of ferritic steel welds". *Science and Technology of Welding and Joining*. Vol. 5. 2000. pp. 81-89.
7. A. J. Smola, B. Schölkopf. "A tutorial on Support Vector Regression". *Neuro COLT2 Technical Report Series*. NC2-TR-1998-030. 1998. pp. 4-18.
8. B. M. del Brío, A. S. Molina. *Redes Neuronales y Sistemas Difusos*. Ed. Alfaomega. Zaragoza. 2001. pp. 76-78.
9. S. Haykin. *Neural Networks. A Comprehensive Foundation*. 2ª ed. Ed. Prentice-Hall. New York. 1994. 1999. pp. 156-255.