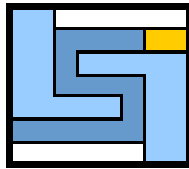


# Proyecto de Tesis

Técnicas de Data Mining para la Diagnósis en Sistemas Dinámicos con Incertidumbre Mediante la Caracterización de Series Temporales.



Departamento de Lenguajes y Sistemas Informáticos.  
Universidad de Sevilla

*Antonio J. Suárez Fábrega*  
*Sevilla, enero 2007*

## **PROYECTO DE TESIS.**

### **Título.**

Técnicas de Data Mining para la Diagnósis en Sistemas Dinámicos con Incertidumbre Mediante la Caracterización de Series Temporales.

### **Alumno.**

Antonio J. Suárez Fábrega

### **Director.**

Rafael Martínez Gasca.

### **Resumen.**

El tema principal del proyecto de tesis incluye la detección y la diagnóstico de fallos. Los fallos, comportamientos anómalos de sistemas y procesos, provocan en los casos más leves efectos económicos, como disminución de la producción de los procesos, debido a paradas indeseadas por mantenimiento o debido a un menor rendimiento, disminución de la calidad de los productos debido al deterioro de los procesos, averías importantes, etc. En casos más graves pueden tener consecuencias que van más allá de las simplemente económicas y ocasionar daños a personas o impactos negativos en el medio ambiente.

La diagnóstico permite a los sistemas de producción actuales mantener altos niveles de seguridad, producción, fiabilidad y calidad. Las técnicas desarrolladas hasta el momento adolecen de altas tasas de falsos positivos y falsos negativos, en especial cuando la diagnóstico es aplicada a sistemas con alta incertidumbre, datos incompletos o incluso cuando únicamente se dispone de información cualitativa.

Respecto de los sistemas con incertidumbre (en los parámetros y debidos al ruido), las técnicas clásicas, tanto las que provienen de la ingeniería de sistemas (técnicas FDI) como las que provienen de la inteligencia artificial (técnicas DX), intentan superar estos problemas con la combinación de técnicas de ambos tipos y buscar caminos intermedios. Aunque el marco de esta tesis podemos considerarla dentro la inteligencia artificial, puesto que usa técnicas de Minería de Datos, usaremos algunas prácticas de FDI que han tenido un más que aceptable grado de éxito.

El proyecto consiste en realizar detección y diagnóstico de fallos en sistemas dinámicos, todo ello de forma eficaz, con coste computacional reducido, de tal forma que pueda realizarse dicha diagnóstico en tiempo real, en presencia de incertidumbre.

## Memoria.

El objetivo de la tesis es desarrollar técnicas que permitan realizar diagnosis automática de sistemas dinámicos con incertidumbre, mediante técnicas de data-mining. Dichas técnicas serán validadas en plantas pilotos y sistemas reales que serán diseñados y monitorizados para tal fin.

Los datos observados serán tratados para conseguir la reducción de información no relevante y resaltar aquella que resulte útil en la tarea propuesta. Para tal fin se tratarán las series temporales provenientes de los sistemas dinámicos de forma cualitativa y cuantitativa, teniendo como primer paso de la tesis la obtención de datos de sistemas reales o plantas pilotos y el tratamiento de estos, el sistema real estará compuesto por un motor DC, un módulo de adquisición de datos y un módulo de control y actuación, y la planta piloto consistirá en un sistema de tanques interconectados, sobre dichos sistemas, se provocarán los fallos de funcionamiento que deberán ser detectados y diagnosticados.

Se implementará una técnica que permita discretizar la información de una forma óptima y automática mediante el uso de clustering, de tal forma que no necesitemos la intervención del experto humano para delimitar el número e intervalo de los niveles cualitativos que tendrá un sistema, y por tanto, puede ser aplicado a cualquier nuevo sistema que se nos presente o a variaciones del actual.

El uso de modelos con parámetros inciertos (intervalares) para detección de fallos ha demostrado su utilidad en diversos casos prácticos. Hay que complementar el sistema de detección de fallos con un sistema de diagnosis, el uso de información incierta (intervalar) nos permitirá que el volumen de datos a manejar sea menor, para ello usaremos técnicas que caractericen las series temporales, mediante índices de similitud cualitativa, Kernels..., así como, una representación de esos datos mediante un sistema de pares ordenados, esto nos evitará la repetición, de tal forma que si tenemos una cadena de símbolos cualitativos que representa a una serie temporal:

$$S_x = \langle s_0, \dots, s_{T-1} \rangle$$

Podremos representarla como un conjunto de pares ordenados, indicando en cada par el valor cualitativo y las repeticiones consecutivas de este.

$$D_x = \langle (s_0, n_0), \dots, (s_p, n_p) \rangle$$

Todo ello nos reduce el número de datos y nos facilita el cálculo de índices de similitud y el uso de Kernels.

Se aplicarán técnicas de aprendizaje supervisado al conjunto de datos observados añadiéndole la información obtenida al caracterizar las series temporales, con el objetivo de identificar los fallos provocados, será parte importante de la tesis el desarrollo de nuevos métodos de detección y diagnosis mediante la combinación de técnicas basadas en modelos y aprendizaje, realizando un análisis de las técnicas de

aprendizaje supervisado con el objetivo de valorar aquellas con un mejor rendimiento (menor número de falsos positivos y falsos negativos). El objetivo es aplicar las técnicas de aprendizaje a la dinámica del residuo generado mediante la comparación de los datos obtenidos del sistema real y del modelo, con objeto de extraer información relevante, susceptible de ser utilizada para la diagnosis del sistema. Esto dará como resultado final una serie de reglas o árboles de decisión que nos permitirán la detección y diagnosis de fallos en tiempo real.

Como resultado final de la tesis obtendremos un sistema de diagnóstico automático que nos permita capturar fallos de tipo abrupto, el sistema de diagnóstico será centralizado y permitirá diagnosticar tanto fallos simples como fallos aditivos, aplicable a un gran abanico de sistemas dinámicos sin la intervención humana, y con la ventaja de utilizar la información del modelo del sistema real cuando se disponga de ella, pero no es una limitación cuando no se dispone del modelo.