

APROXIMACIÓN AL DIAGNÓSTICO ESTADÍSTICO DE AVERÍAS MEDIANTE LA UTILIZACIÓN DE VARIABLES MULTIDIMENSIONALES

LARIO ESTEBAN, F. C.

El objeto de este trabajo es plantear las posibilidades de utilización del Análisis de Datos y el Reconocimiento de Formas a un concepto de Entretenimiento preventivo que amplie los conceptos clásicos a través del análisis Causa-Efecto. Mediante el Análisis de Datos y los conceptos de Problemas de Clasificación o Reconocimiento de Formas, puede plantearse el conocimiento de los Estados de Degradación de un Equipo o Sistema; las técnicas de Comprensión de Datos y el Análisis Factorial de Correspondencias pueden resolver la fase de clasificación, para posteriormente mediante la Asignación (o Identificación) de los individuos a las clases automatizar el Diagnóstico de Averías. De nuevo aquí se utiliza como Identificador el basado en el estimador de k-NN o k-Nearest Neighbor (o k-vecino más próximo). Este planteamiento permitiría integrar la experiencia y conocimientos históricos de la evolución del sistema gracias a la utilización de Muestras de Aprendizaje, que precisamente, por ello es supervisada con Profesor.

1. INTRODUCCION

La idea clásica del Control Estadístico de Calidad en la Fabricación, utilizando variables unidimensionales, se centra sobre una característica del producto fabricado, y según las variaciones de dicha característica, se infiere el funcionamiento adecuado o no del proceso de fabricación. Según como, podría considerarse como un Control de Recepción a la salida de la máquina, y en lugar de considerar rechazable el lote producido y devolverlo, lo que se hace es parar la máquina y ajustarla.

Si lo que realmente interesa es conocer el estado de funcionamiento del proceso de fabricación, si lo que se pretende es conocer el estado de degradación del equipo, y las causas que lo originan, parece natural que se trabaje con variables multidimensionales que permiten integrar las distintas magnitudes que caracterizan el proceso y el producto, de acuerdo con lo que se expone en /8/.

Si se es capaz, una vez conocido el valor que toma la variable multidimensional, obtiene mediante sensores o captadores situados en distintos puntos del proceso, de asignarla a una de las Clases o Estados en que se ha dividido la evolución (o degradación) del

Equipo, se habrá efectuado un Diagnóstico Estadístico del Estado de Degradación (o Avería) en que se encuentra el sistema en el momento considerado.

En definitiva se está planteando un Problema de Identificación o de Reconocimiento de Formas, con todas sus fases tal y como expone Pau /11/. Primero se tendrán que definir, e incluso explicitar, los tipos de Estados de Degradación del Sistema que se consideran, o sea se plantea la Partición o Clasificación; posteriormente y según la Formas, variables multidimensional cuyas componentes definen características pero no el Estado del Equipo, que se obtengan en cada momento asignar el Equipo, Proceso o Sistema a una de las Clases (o Estados) de Degradación definidos, con lo cual se resuelve la fase de Identificación o Asignación.

Es en la fase previa a la Clasificación y en ésta propiamente dicha donde tiene sentido la utilización del Análisis de Datos, como puede apreciarse en Pau /10 y 12/ y en Lario /8/. Como Operador de Reducción, y a partir de una muestra de Aprendizaje (con o sin profesor), permitirá disminuir la dimensión de la variable, de manera que con un mínimo de dimensión se alcance un máximo de la información mantenida en la variable; como opera--

Francisco Cruz Lario Esteban. E.T.S. Ingenieros Industriales de Valencia. Cátedra de Organización de la Producción. Camino de Vera, s/n. Valencia

Article rebut al març 81.

dor de Partición y actuando sobre la Muestra de Aprendizaje de dimensión reducida, permitirá la definición explícita de las Clases o Estados de Degradación del Equipo.

Por lo tanto en el problema del Diagnóstico de Averías se manifiesta el esquema integrado de un Problema de Clasificación o Reconocimiento de Formas, incluyendo tanto la Reducción de la dimensión, la Partición de las Formas de la Muestra de Aprendizaje, como la identificación de la Forma, y con ella la determinación del Estado del Proceso cuando se ha obtenido la forma. Debe hacerse constar que este sistema de análisis integra la experiencia histórica acumulada mediante lo que aquí se ha llamado Muestra de Aprendizaje, y trata de auxiliar en el difícil tema del Diagnóstico de Averías, incluso abriendo la posibilidad del Diagnóstico automático de Averías.

En definitiva el Diagnóstico de Averías trata de resolver dos problemas prácticos que la Teoría de la Fiabilidad sugiere, pero no resuelve, la Detección y Localización de la Avería.

En un sentido amplio y real del Diagnóstico de Averías debería hablarse del Diagnóstico de Anomalías de Funcionamiento de un Sistema o Equipo. La noción de Anomalía de funcionamiento no tiene más sentido que con referencia a unas condiciones de empleo del Sistema, de aquí que en ocasiones se utiliza la terminología Estados de Degradación del Sistema, -entre los cuales y con consecuencias importantes-, se halla el paro o interrupción del funcionamiento del Sistema, generalmente, llamado Avería. En efecto, bajo la óptica de vigilancia continua del funcionamiento de un Sistema, es insuficiente caracterizar un estado de Avería mediante la superación de un techo o límite de un parámetro de funcionamiento o mantenimiento. La Avería física puede resultar de la aparición conjunta de desviaciones de varios parámetros con respecto a sus valores iniciales; al estudio, análisis y clasificación de los distintos parámetros, generalmente expresados en forma de vectores, permite la vigilancia y determinación de los Estados de Degradación del Sistema.

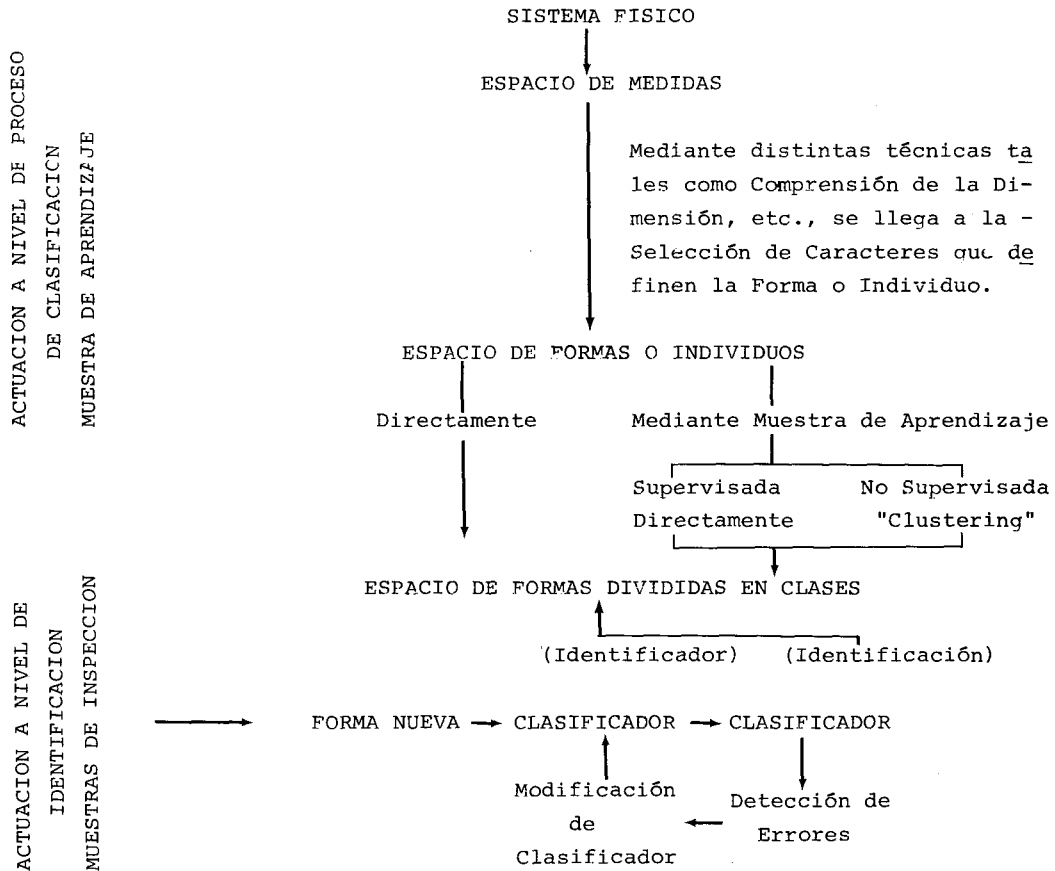


Fig. 1: Esquema integrado del problema de clasificación

Al hablar de Diagnóstico de Averías debe diferenciarse entre los conceptos de Diagnóstico Lógico de Averías, según Coutensou /2/ y Diagnóstico Estadístico de Averías y Automatización del Diagnóstico Estadístico, tal y como aparece en Pau /16/. El Diagnóstico lógico, del cual no se ocupará este trabajo, trata de Sistemas individualizados sobre los que se efectúan directamente test de localización de Averías sin tener en cuenta informaciones a priori. Por el contrario el Diagnóstico Estadístico de Averías conduce al análisis de un conjunto de sistemas idénticos cuyos síntomas de comportamiento se observan a posteriori mediante la recolección sistemática de informaciones de todo orden.

Con otras palabras, puede entenderse como finalidad del Diagnóstico Estadístico de Averías el reconocimiento, en una lista fijada de Síntomas-Causas-Componentes, de las Causas de Averías y las Averías según los valores de las observaciones realizadas sobre determinadas variables (sin que ello implique un Desmontaje previo), valores que se compararán con la experiencia acumulada sobre la población de Sistemas idénticos. Se deduce pues que el Diagnóstico Estadístico de Averías exige un Sistema de Recogida y Codificación de información.

El fin de todo procedimiento de Diagnóstico, es pues, identificar una Avería o Conjunto de Averías, sobre la base de:

- Informaciones referidas al pasado del Sistema o Equipo, comprendidas las Operaciones de Mantenimiento y su utilización en el proceso. Informaciones que constituyen lo que ha dado en llamarse Datos de Aprendizaje (almacenados generalmente en un Banco de Datos de Fiabilidad y Mantenimiento).
- Informaciones sobre las circunstancias de los Estados del Sistema, mediante valores de observación o señales, valores que por ser generalmente multidimensionales constituyen las coordenadas de un vector llamado Forma Vectorial.

De manera que una investigación de las "correlaciones" entre la Forma Vectorial (expresión analítica de síntomas) y los datos de aprendizaje, permita clasificar el Equipo o Sistema averiado en uno de los posibles ti-

pos de causa de avería. La decisión final en cuanto a la clasificación de la Forma Vectorial en uno de los tipos de Estados de Degradación o Avería, se realizará mediante una Regla de Decisión fijada, de modo similar a como se asignaba una Pieza a la Clase Aceptable o Rechazable en Control Estadístico de Calidad multidimensional /8,9/.

De lo dicho anteriormente sobre el Análisis de Datos, pueden deducirse de entre sus objetivos el tratamiento de informaciones estadísticas multidimensionales, -entre las cuales existen correlaciones-, estando particularmente destinado a poner en evidencia las Relaciones de Causalidad susceptibles de explicar el fenómeno observado y analizado.

Por lo tanto los Métodos del Análisis de Datos pueden conducir a una explicitación y visualización fiel de las relaciones causales estimadas en los Datos de Aprendizaje, lo que facilitará las labores del Diagnóstico Estadístico de Averías.

1.1 RECOGIDA DE DATOS

Al definir el Diagnóstico Estadístico de Averías se comentó la recolección o Recogida Sistemática de Información (de Datos) de todo orden, a partir de la cual se realiza el diagnóstico del Estado de Degradación del Sistema. A consecuencia de ello y del papel fundamental jugado por dichas informaciones a priori, datos iniciales o Datos de Aprendizaje, tiene una gran importancia el Estudio y Diseño del Sistema de Recogida de Información para el establecimiento de la Fiabilidad y Mantenimiento de los Equipos o Sistemas; es lo que en la literatura técnica correspondiente recibe el nombre de Banco de Datos de Fiabilidad y Mantenimiento.

El objetivo del citado estudio será definir un conjunto lo suficientemente amplio de parámetros aptos para representar la experiencia adquirida sobre cada uno de los componentes o subsistemas del Equipo, objetivo que exigirá un conocimiento real y práctico del citado Equipo, ya que no puede plantearse el conjunto de parámetros de una forma teórica o empírica, sino todo lo contrario; lo cual no priva de una metodología del Diseño de Recogida de Datos.

Dentro del capítulo de recogida de la experiencia adquirida, pasada o presente, tiene un papel determinante el conocimiento de la documentación que relativa al Equipo objeto de estudio pueda encontrarse en los distintos centros de documentación e información, servicios oficiales constructores o suministradores de Equipos, utilizadores, etc. -- Igualmente se planteará el problema de unificar las distintas informaciones y los distintos soportes de información referentes a Datos de Fiabilidad y Mantenimiento de un Equipo o sistema.

Resuelto el estudio y Diseño del Banco de Datos de Aprendizaje, se tratará de escribirlos de manera que permite su análisis mediante las técnicas del Análisis de Datos, para lo cual se seguirán los conceptos y notaciones utilizadas por Pau /15/

Supóngase reunidas y conocidas todas las informaciones referentes a los individuos i , $i \in I$, referidas a las intervenciones u observaciones j , $j \in J$, obtenidas sobre todas las partes o subsistemas del Equipo que se han definido en el Diseño del Banco de Datos de Fiabilidad y Mantenimiento.

En un momento dado se definen los Datos de Aprendizaje sobre los componentes o partes del Equipo, como el conjunto $k(I,J) = \{k(i,j), i \in I, j \in J\}$ de datos recogidos hasta este momento, de manera que:

$$j(i,j) = \begin{cases} b(i,j) = & \left. \begin{array}{l} =1, \text{ cuando el sistema se encuentra en el Estado } i, \text{ en la observación } j. \\ =0, \text{ el sistema no se encuentra en el Estado } i, \text{ en el instante u observación } j. \end{array} \right\} \\ a(i,j) = & \text{representa el valor de la magnitud } i \text{ en el instante } u \text{ observación } j. \end{cases}$$

según los tipos de equipo o procesos a controlar, y en consecuencia los parámetros utilizados para definir cada Estado del Sistema o Equipo, los datos recogidos serán del tipo $b(i,j)$ o del $a(i,j)$, integrando lo que se conoce como Tabla Explícita $k(I,J)$ ya que sobre ella se encuentran reunidas de forma explícita todas las informaciones sobre los individuos $i \in I$ tomadas en cada instante u observación $j \in J$.

En algunos casos las especiales características del Equipo hace que se presente al mismo tiempo datos en forma $a(i,j)$ y $b(i,j)$, resultando entonces interesante la construcción de unas Tablas de Datos especiales.

Recogido un conjunto de Datos $k(I,J)$ en forma de Tabla Explícita donde haya términos del tipo $a(i,j)$ y $b(i,j)$ se define $n(i,j)$, para el caso de que varios i sean de carácter alternativo, entonces

$$n(i,j^*) = \begin{cases} - \text{ Si } i \text{ no tiene ninguna alternativa, o sea toma un valor único } n(i,j^*) = \text{valor medio } k(i,j). \\ - \text{ Si } i \text{ tiene alternativas, entonces } n(i,j^*) = \text{número de intervenciones en las que } k(i,j) = 1 \end{cases}$$

Otras veces tiene interés presentar la Tabla en forma de Frecuencias relativas $k^*(i,j^*)$ de manera que

$$k^*(i,j^*) = \begin{cases} - \text{ Si } i_k \text{ no tiene alternativas} \\ \quad k^*(i,j^*) = n(i,j^*) \\ - \text{ Si } i_k \text{ tiene alternativas,} \\ \quad k^*(i,j^*) = \frac{n(i,j^*)}{\sum_k n(i_k,j^*)} \end{cases}$$

En lo que se refiere al resto de este trabajo se han tenido sobre todo en cuenta las Tablas Explícitas del tipo $a(i,j)$, correspondientes a Correspondencias Estadísticas, a partir de las cuales y tal como se ha indicado pueden obtenerse tablas de Correspondencias Aleatorias del tipo $p(I,J) = \{p(i,j), i \in I, j \in J\}$ mientras que las Tablas de correspondencias del tipo $b(i,j)$, no han sido objeto de estudio y utilización preferente, si bien para estas últimas también pueden definirse su transformación en Correspondencia Aleatoria mediante

$$\forall (i,j) \in I \times J, \quad p_b(i,j) = \frac{a(i,j)}{\sum_{(i,j) \in (I,J)} A(i,j)}$$

1.2 COMPRESIÓN Y RELACIONES DE CAUSALIDAD

Al hablar del Análisis de Datos como operador de Compresión, puede considerarse dos técnicas distintas, actuando sobre una Tabla de Números $p(i,j)$ que define la correspondencia aleatoria estudiada (la Tabla recibe el nombre de Tabla de Contingencia)-, el

Análisis de Componentes Principales y el Análisis Factorial de Correspondencias, /1/.

Al estudiar el Análisis en Componentes Principales se observa la posibilidad de reducir el número de observaciones o variables dentro de cada individuo, utilizando una transformación sobre las observaciones de partida de manera que se obtenga un número inferior de observaciones o variables dentro de cada individuo. El Análisis en Componentes Principales permitirá seleccionar el número de observaciones sintéticas al mismo tiempo que define los primeros componentes principales que encierran una proporción importante de la varianza total de los individuos; permitiendo ver cual es la dimensión verdadera del individuo, cuales las importancias respectivas λ_j de las observaciones sintéticas de los individuos, ligadas a las primeras componentes principales, y por último permitir la representación de todo nuevo individuo en el espacio de observaciones variables sintéticas (espacio de dimensión inferior al de observaciones iniciales).

No obstante en lo que se refiere a la representación y sobre todo a las relaciones de Causalidad, el Análisis de Componentes Principales no permite relacionar variables (u observaciones) e individuos, sino tan sólo analizar lo que ocurre entre las observaciones de un lado y los individuos de otro. Dicho de otra manera, en el caso que se trata de estudiar no se podrán deducir o diagnosticar sobre los Equipos (individuos) por referencia directa a las observaciones.

Por el contrario el Análisis Factorial de Correspondencias es un método numérico de Comprensión de Datos, presentados bajo forma de una Correspondencia aleatoria, y cuyo objeto es tratar informaciones estadísticas de todo orden entre las cuales hay un gran número de correlaciones, poniendo en evidencia las relaciones de causalidad.

Mediante el A.F. de C. se determinan los factores en función de los cuales es posible interpretar objetivamente la mayor parte de las relaciones causales agrupando en una misma representación gráfica las observaciones y los individuos (Equipos o Instalaciones) - teniendo en cuenta sus relaciones con los factores puestos en evidencia mediante el A.F. de C., como puede observarse en /10 y 12/

En definitiva el Análisis de Correspondencias puede utilizarse como técnica matemática para la Síntesis de información relativa al Mantenimiento de un Sistema, sistema representado por una Correspondencia Aleatoria $p(i,j)$, $i \in I$, $j \in J$, y en distintas fases tales como:

- Fase de compresión de Datos
Permite reducir la dimensión del individuo cuando la menor dimensión de la Tabla de correspondencia aleatoria es demasiado grande, sin que sea necesaria ninguna clasificación a priori de los elementos del Conjunto de Individuos o de Observaciones
- Fase de aprendizaje
La representación gráfica sobre planos factoriales de observaciones e individuos provee de una clasificación previa de los elementos, basada en proximidades geométricas.
La representación en el Espacio de las observaciones permite detectar entre ellas las observaciones redundantes, y eliminarlas en el conjunto de Datos de Aprendizaje por simple reducción de las observaciones tomadas en cuenta.
- Fase de Diagnóstico y Reconocimiento de Formas
Dados los Datos de Aprendizaje en forma de Tabla explícita el A.F. de C. tiene la ventaja de poner en evidencia e interpretar, con riesgos de error pequeños, las relaciones de Causalidad entre Individuos (Equipos) y Observaciones. De otra parte este método permite visualizar simultáneamente sobre unos mismos gráficos los Equipos y los puntos representando en media cada una de las observaciones sobre éstos cosa que no es posible utilizando otros métodos de Análisis de Datos.
Finalizado el Aprendizaje y la Comprensión de los Datos de Aprendizaje, el Operador de Comprensión, da mediante una simple fórmula lineal las coordenadas de un nuevo equipo desconocido en el Espacio comprimido de dimensión r .

1.2.1 Utilización del Análisis de Correspondencias

Al hablar de los conceptos de Avería y Diagnóstico se ha caracterizado la existencia de

un Problema de Asignación, -la de un vector representativo de la situación instantánea - de un Sistema o Equipo a uno de los Estados de Degradación del citado Equipo o Sistema-, problema que exige la existencia de un Conjunto de Datos (o Muestra de Aprendizaje) -- que puede o no estar, ya, dividida en los -- distintos estados o Clases de Degradación -- del Sistema; en definitiva en la aplicación del Reconocimiento de Formas al Diagnóstico de Averías /13, 14/.

Si dicha Muestra de Aprendizaje estuviese ya particionada en las distintas Clases, nos en contraríamos ya directamente con la asignación o Clasificación de un nuevo individuo - entre las Clases o Estados de Degradación -- del Sistema, en definitiva en lo que sería - un Diagnóstico Estadístico Automático de Averías.

Sin embargo puede ocurrir y de hecho ocurre, que la Muestra de Aprendizaje no está particionada, encontrándose entonces el problema en la fase previa. Se tienen distintas representaciones de los Estados del Sistema, mediante individuos cada uno de ellos caracterizado por varias observaciones o variables, siendo necesario el establecimiento de relaciones de Causalidad entre individuos y observaciones. Es precisamente para plantear estas relaciones, e incluso obtener particiones gráficas de los individuos y observaciones, que se utiliza el A. Factorial de Correspondencias, que igualmente actúa como un Operador de Reducción del Espacio de las observaciones.

De igual modo que con carácter previo al problema del Diagnóstico en sí ha tenido que plantearse el de Recogida de Datos, un vez realizadas las operaciones de A.F. de C. se hace necesaria una Interpretación de los resultados obtenidos, interpretación que basada en el juicio de experiencia del fenómeno analizado, se apoya sobre algunos Test de Hipótesis.

2. AUTOMATIZACION DEL DIAGNOSTICO DE AVERIAS

Al hablar de la Teoría General de Averías ligada al Mantenimiento y la Fiabilidad se delimitó con toda claridad la finalidad del Apartado del Diagnóstico de Averías, la Detec

ción y la Localización de la Avería; al tratarse de un Diagnóstico Estadístico las técnicas utilizadas se basarán exclusivamente - en el Análisis Estadístico de todas las informaciones disponibles sobre el funcionamiento y el entorno del Sistema o Equipo hasta el momento de la Avería.

De lo que se tratará en adelante es de establecer unas técnicas que permitan la automatización del citado análisis, a fin de que - el desmontaje y verificación de distintas partes del Sistema no tenga que ser realizado más que como último recurso de Localización o bien en el Estudio de Identificación de la causa de Avería, de manera que habiéndose determinado automáticamente la o las causas de avería, se tratará únicamente de desmontar la o las partes del sub-sistema -- probablemente averiados.

Técnicamente el Diagnóstico Automático de Averías de un Sistema dado es realizable mediante un mini-calculador de programa fijo, o bien mediante una unidad aritmética funcionando a tiempo partido asociada a un Equipo.

Se dirá pues que el Diagnóstico automatizado consiste en reconocer las causas de Avería - (o de los Estados de Degradación) a partir de observaciones realizadas sobre señales y captos del propio sistema. Si los valores de las distintas observaciones tomadas sobre el Sistema pueden representarse mediante magnitudes medibles representadas como componentes de un vector multidimensional, estas observaciones constituyen lo que se denomina - Formas, y el problema del Diagnóstico de Estados de Degradación o Averías queda asimilado a un problema de Reconocimiento de Formas

En otras palabras formular un Diagnóstico Automatizado quedará circunscrito a tomar la decisión de reconocer la Forma-Observación - dentro de las Formas Dato que definen cada uno de los Estados de Degradación del Sistema.

Debe hacerse constar que el Reconocimiento de Formas, asignando una Forma-Observación a una de las Clases o Estados de Degradación, implica con carácter previo la existencia de un Aprendizaje, aprendizaje que desde el punto de vista de las técnicas de Análisis se manifiesta por los Datos de Aprendizaje, da

tos que podrán venir o no comprimidos en --- cuanto a su dimensión (actuación de un Operador de compresión tipo Análisis de Componentes principales o mejor Análisis Factorial - de Correspondencias), y que podrán estar ya previamente clasificados en los distintos Estados de Degradación del Sistema (Datos de Aprendizaje Supervisado con Profesor) o al no estarlo tendrán que sufrir una Clasificación previa (Datos de Aprendizaje no supervisados sin Profesor), que podrá realizarse con carácter aproximativo y en un gráfico de dimensiones inferiores a las reales mediante un Análisis Factorial de Correspondencias, o con un carácter más específico y riguroso utilizando algoritmos de "Taxonomía" /3/.

Llegados a este punto puede establecerse una relación entre el Control de Calidad de Fabricación medido sobre el artículo sometido a los procesos de fabricación del Equipo, y el Mantenimiento del citado Equipo productivo. Puede pensarse que los parámetros de calidad del artículo fabricado son observaciones o medidas realizadas sobre el Equipo en instantes dados (individuo), de manera que - el Control de Calidad de Fabricación proporciona al mismo tiempo observaciones que mediante el Diagnóstico Automático de Averías permiten informar del Estado de Degradación (mantenimiento) en el que se encuentra el -- Sistema o Equipo.

Como caso práctico referido a lo anterior y con marcadas aplicaciones de tipo industrial podría hablarse de los Regularímetros Upster de cada vez mayor utilización en la Industria Textil en su rama Hilatura y que mediante la utilización del parámetro Regularidad del Hilo es capaz al mismo tiempo de definir la Calidad del Hilado (en cuanto a su regularidad) y la causa previsible del anómalo funcionamiento del Equipo productor del citado - hilo o mecha.

2.1 DATOS DE APRENDIZAJE, SU TRATAMIENTO

Los datos de Aprendizaje se presentarán en - forma de una Tabla Explícita $\{k(i,j), i \in I, j \in J; k(i,j) \geq 0\}$ obtenida situando todas -- las informaciones u observaciones $j \in J$ en cada uno de los instantes o individuos del Equipo. Cada uno de los $k(i,j)$ representará - la expresión analítica de la experiencia histórica de un instante o vector individuo del

Equipo, de manera que la experiencia necesaria para construir los Datos de aprendizaje estará dada generalmente por un gran número de $k(i, \cdot)$, tal y como queda expuesto en /15, 8/.

El conjunto de $k(i, \cdot)$ podrá presentarse de - dos formas:

- Los individuos de $k(i, \cdot)$ que integran la Muestra de Aprendizaje están previamente agrupados en Clases, definiéndose entonces la Muestra del Aprendizaje como Supervisada con Profesor.
- Los individuos $k(i, \cdot)$ que integran la --- Muestra de Aprendizaje no han sido previamente agrupados en Clases, definiéndose - entonces la Muestra de Aprendizaje como - no-supervisada sin Profesor.

Incluso en el caso de Muestras Supervisadas con Profesor se necesita explicitar analíticamente la expresión de cada una de las Clases en que se ha particionado el Conjunto de Formas (aquí podría utilizarse el Análisis - Discriminante, como auxiliar en los Problemas de Clasificación e Identificación; si -- bien aplicado a individuos que sigan una Ley Normal Multidimensional).

Por el contrario en el caso de Aprendizaje - no-Supervisado es necesario como primer paso resolver el problema de Clasificación previa de los individuos $k(i, \cdot)$. Para ello pueden tilizarse dos caminos, los Métodos basados - en Algoritmos de Taxonomía que utilizan fundamentalmente funciones de semilaridad expresándose con la sola ayuda de las observaciones o componentes de los $k(i, \cdot)$ y una estimación del número máximo de clases, o los Métodos Estadísticos del tipo Análisis Factorial de Correspondencia, Análisis en Componentes Principales, etc., que utilizan técnicas de test estadísticas para medir las características de similaridad y no hacen ninguna hipótesis previa sobre el número de Clases.

En lo que a tratamiento matemático (e incluso informático) de lo anterior se refiere, - la dimensión de los Datos de Aprendizaje --- coinciden con las del individuo $k(i, \cdot)$ a clasificar automáticamente, y por lo tanto alcanza un valor entero suficientemente grande tal que puede acarrarar dificultades, de --- aquí la idea de Compresión de Datos.

Se ha hablado de que la Automatización del Diagnóstico podría sugerir la implantación de un micro-procesador en el Equipo que realice las funciones propias de la Clasificación automática; si dicho micro-procesador tuviese que clasificar y/o discriminar los individuos $k(i, \cdot)$ de la Muestra de Aprendizaje, al mismo tiempo que comprime la dimensión del Espacio de Observaciones, dejaría de ser un micro-procesador de aquí que en lugar de trabajar sobre individuos $k(i, \cdot)$ de la Muestra de Aprendizaje lo hará sobre los de dimensión comprimida, al igual que ocurrirá con el $k(i, \cdot)$ inclasificado, que transformará a su dimensión comprimida.

2.2 LA ASIGNACION DE LA FORMA A UNA CLASE

Una vez realizada la Comprensión de Datos (mediante el A.F. de Correspondencias u otra técnica de compresión) sobre los datos de la Muestra de Aprendizaje, el Reconocimiento de la Forma puede descomponerse en tres fases:

- La discriminación, que consiste en la elección del método analítico de separación de las Clases entre sí, y la adopción de unas medidas de similitud entre Forma y Clase de la Muestra de Aprendizaje.
- La Decisión, que consiste en aplicar un criterio de Decisión, basado en la utilización de la función Discriminante /7/, o en la del Identificador del k-vecino más próximo (k-Nearest-Neighbor o k-NN), tiene por finalidad clasificar la Forma (individuo observado $k(i, \cdot)$) en una de las Clases de Aprendizaje que representan uno de los Estados de Degradación del Equipo. Precisamente este criterio de Decisión se adopta sobre las medias de similitud definidas en la fase anterior.
- En una fase posterior con sentido práctico y real, la visualización de la decisión tomada y la posibilidad de ajustar un sistema de regulación que permita la desconexión o aviso automático del Estado de Degradación en que se encuentra el Equipo o Sistema.

2.2.1 Utilización de la técnica del k-NN en la Asignación

Como en cualquier proceso de Diagnóstico del tipo estudiado hasta ahora, la aplicación de la técnica del k-NN, -descrita en /4/, y ampliada en cuanto a Optimización del Estimador de la Densidad y del Riesgo a Bayes en la Identificación mediante el k-NN, en /5 y 6/, no priva de las fases previas de Recogida de Datos, con determinación de la Muestra de Aprendizaje (Supervisada o no), seguida de la fase optativa de la compresión de Datos y finalmente la fase de Asignación o Clasificación de los distintos individuos $k(i, \cdot)$ no pertenecientes a la Muestra de Aprendizaje que se desean diagnosticar automáticamente.

Conocidos ya los Datos de Aprendizaje, y presentados en forma de una tabla explícita $\{k(I, J) = k(i, j) \geq 0, i \in I, j \in J\}$ obtenida recogiendo todas las informaciones sobre las $j \in J$ observaciones de cada individuo o situación instantánea del Equipo o Sistema $i \in I$ (o bien sobre distintos Equipos todos ellos idénticos) se supone la actuación de un Profesor que supervise la Muestra de Aprendizaje de manera que cada individuo de la Muestra de Aprendizaje esté asignado a una Clase representada aquí por w_1 , de manera que cada Clase se define por el Conjunto de individuos que la integran.

Se supone que el número de Clases de Estados de Degradación del Equipo 1, será relativamente pequeña comparado con el número de Individuos de la Muestra de Aprendizaje, dada la problemática de la estimación de los Riesgos de mala Identificación cuando el número de clases es elevado.

Si se tiene en cuenta que en el Criterio de Decisión, -tanto sea mediante la función Discriminante como a través del k-NN-, intervendrá la probabilidad a priori del Estado de Degradación w_1 , se considerará conocida la Distribución de dicha Probabilidad, bien a partir de una Estimación sobre la muestra de Aprendizaje, bien por especificaciones de otro tipo.

Si se ha realizado la Comprensión de Datos, generalmente mediante el Análisis Factorial de Correspondencias /15,1/, los individuos de la Muestra de Aprendizaje, se representa-

rán en un Espacio de dimensión r , inferior a la del Espacio de los Individuos y a la del Espacio de las Observaciones.

Si el Sistema de Diagnóstico Automático se realiza mediante Muestreo Periódico de la situación del Equipo, periódicamente se tendrá un individuo $k'(i', \cdot)$, con la información correspondiente, $k'(i', j)$, de todas sus variables o componentes que tendrá que ser asignado a una de las Clases o Estados de Degradación del Sistema, con lo que periódicamente se tendrá la situación del Equipo con respecto a su propio Mantenimiento.

Se considerará el $k'(i', \cdot)$ como una nueva forma perteneciente a una de las clases w_1 (desconocida).

Llegados aquí se aplicará la regla del vecino más próximo (k-NN), si bien situando los distintos individuos de la Muestra de Aprendizaje y el individuo a Clasificar en un Espacio r -dimensional, siendo r la dimensión del Espacio comprimido de individuos.

3. CONCLUSIONES

Se ha tratado de revisar el Diagnóstico Estadístico de Averías mediante la utilización de variables multidimensionales que de una parte permitan el Análisis Causa-Efecto en la evolución de un Sistema (Equipo), y que de otra al definir unos vectores de estado (variable multidimensional en momentos sucesivos) que reflejan la Situación del Sistema permitan la Identificación de la Fase de Degradación o Avería en que se encuentra el Equipo o Sistema.

Será a través del Análisis de Datos, actuando como Operador de Reducción o Compresión, y utilizando el Análisis de los Componentes Principales y de las Correspondencias Factoriales, que pueden establecerse relaciones Causa-Efecto. Si interesa un Diagnóstico automático de la fase en que se halla el Equipo en un momento dado, mediante el Reconocimiento de Formas o Patrones y una vez resuelta la partición en Clases o Clasificación, se podrá conocer el Estado en que se encuentra el Equipo gracias a la Asignación o Identificación del Vector de Estado en ese momento a través de un Identificador, sea és

te el de la Función Discriminante o el basado en el estimador del k-NN.

Gracias a la elaboración teórica e instrumental del Diagnóstico Estadístico de Averías, puede llegar a una concepción más global y ajustada a la realidad del tema del Control de Calidad en la Industria, tanto el referido a las Materias Primas o de Recepción como al de Fabricación, ligándolo al Mantenimiento, las Averías y su Diagnóstico e incluso a la Fiabilidad, tanto conceptualmente como a través de las técnicas e instrumentos matemáticos y estadísticos utilizados.

La idea clásica del Control de Fabricación utilizando variables unidimensionales se centra sobre una sola característica, y además, del Producto elaborado, mientras que la utilización de variables multidimensionales permite considerar el Proceso de Fabricación como tal, sobre el cual y mediante sensores o captadores pueden conocerse los valores que distintas variables del Proceso e incluso del Producto van tomando en distintos momentos.

Se podrá pasar pues de un Concepto de Calidad de Fabricación mediante control de una variable del Producto, a una calidad de Fabricación sobre el Proceso de Fabricación como tal y gracias a la variable multidimensional que recoge información sobre el Producto y del Proceso, en definitiva se está actuando con mayor propiedad sobre la idea del Control de Fabricación.

En definitiva, si en el Control de Fabricación se emplea la del Producto, no es tanto para aceptar o rechazar el Lote de Fabricación, -que sería un Control de Recepción a salida de un Proceso-, cuando para conocer el desajuste (Estado de Degradación del Proceso o Sistema) y poder reglar y ajustar el Proceso que es quien define realmente la Calidad de Fabricación.

Llegados aquí se ve la auténtica posibilidad de relacionar estrechamente las ideas de Mantenimiento, Diagnóstico de las Fases de Degradación del Sistema con la Calidad de Fabricación, entendida como proceso operativo en el que hay parámetros del propio Proceso y de los resultados parciales (semi-elaborados) o totales (Productos acabados).

Debe destacarse igualmente, y desde otra --- perspectiva, los interesantes resultados que cara a los temas de Control, -tanto de Pro-- ducción y Calidad-, Mantenimiento y Diagnòs-- tico de Averías, han podido obtenerse gra--- cias a la forma específica en que se han ma-- nejado grandes cantidades de Información acu-- mulada en Bancos de Datos (o muestras de A-- prendizaje, Supervisadas o no) y referidas a Sistemas Complejos.

Indiscutiblemente la dificultad y compleji-- dad del Diagnòstico, tanto en la fase de co-- nocimiento de las relaciones Causa-Efecto co-- mo en la Automatización del Diagnòstico de - la Fase de Degradación, no radica exclusiva-- mente en la solución del modelo planteado, - sino como en toda metodología científica, in-- cluye el contraste con la realidad y el cono-- cimiento práctico. De aquí, en el caso de - aplicación Industrial, la importancia de la Muestra de Aprendizaje supervisada con el -- Profesor, que al mismo tiempo que se consti-- tuye en Banco de Datos para la solución del Modelo, debe reservarse una parte para el -- contraste y verificación de los resultados. De todo lo anterior se deduce la importancia que para cada caso práctico tendrá toda la - información acumulada de carácter práctico y pragmático que puede ayudar tanto en el Plan-- teamiento del Modelo como en la verificación de sus soluciones.

En cuanto a campos de trabajo abiertos ca--- brían los referentes a la Estimación de Ries-- gos y Errores de Clasificación en el caso de que existen más de dos clases, -tanto desde el punto de vista general, como aplicados al caso de la Distribución condicional Normal - multidimensional, como cuando se utilicen es-- timadores paramétricos de la densidad de dis-- tribución del tipo k-NN-, pero ajustando ya los conceptos teóricos al problema y circuns-- tancias específicas y concretas del caso.

4. BIBLIOGRAFIA

/1/ BENZECRI, J.P. et altres: "L'Analyse -- des données", Tomos I y II, Dunod, París 1978.

- /2/ CONTENTSOU, J.N.: "La Logique du Diagnos-- tic", ENSAE, Toulouse, 1977 (Polico---- piées).
- /3/ ESCUDERO, L.F.: "Reconocimiento de Pa-- trones", Paraninfo, Madrid, 1977.
- /4/ FUKUNAGA, K.: "Introduction to statisti-- cal Pattern Recognitions", Academic --- Press Inc., N.Y., 1972.
- /5/ FUKUNAGA - HOSTETLER: "Optimization of k-Nearest-Neighbor Density Estimates --- I.E.E.", Transactions on Information Theo-- ry, Vol. IT-19, pgs. 320 a 326, May 1973 No 2.
- /6/ FUKUNAGA - HOSTETLER: "k-Nearest-Neigh-- bor Bayes-Risch Estimation", I.E.E., --- Transactions on Information Theory, vol. IT-21, pgs. 285 a 293, May 1975, no 3.
- /7/ LACHENBRUCH, P.A.: "Discriminant Analy-- sis", Hafner Press, N.Y., 1975.
- /8/ LARIO ESTEBAN, Fco.C.: "Métodos Esta-- dísticos en el Control de Calidad Mul-- tidimensional por medidas y en el Diag-- nòstico de Averías. Estimación y Análi-- sis de Datos", Tesis doctoral inédita, Barcelona, Mayo 1979.
- /9/ LARIO ESTEBAN, Fco.C.: "Aproximación - al Control Estadístico de Calidad me-- diante variables multidimensionales", Qüestiió, V.4, no3, pgs. 119 a 136, -- Septiembre 1980.
- /10/ PAU, L.F.: "Diagnosticque Statistique. Sythese des informations relatifs à la maintenance d'un material aeronauti-- que", L'Aeronautique et l'Astronauti-- que, no 34, pgs. 69 a 76, 1972.
- /11/ PAU, L.F.: "Applications de la Reconais-- sance des formes dans le domaine aeros-- patial" L'Aeronautique et l'Astronauti-- que no 42, 1973-4, págs. 63 a 78.
- /12/ PAU, L.F.: "Diagnostic statistique et sythese de informations relatives à la fiabilité et à la maintenance d'un material électromecanique", L'Onda é-- lectrique, vol. 54, no 10, pgs. 529 a 537, Decembre 1974.

- /13/ PAU, L.F.: "Applications of Pattern recognition to the Diagnosis of Equipment Failures", Pattern Recognition J. vol. 6, pp. 3-11, 1974.
- /14/ PAU, L.F.: "Diagnóstico de Averías de Equipos mediante el Reconocimiento de Formas", IEEE, Transactions on Reliability, vol. R-23 n°3, pgs. 202-208, August 1974.
- /15/ PAU, L.F.: "Diagnostic des Pannes dans les Systèmes. Approche par la Reconnaissance des Formes", CEPADUES, Editions Toulouse, 1975.
- /16/ PAU, L.F.: "Automatisation du Diagnostic des Pannes. Analyse des Correspondances bayessiennes", Publicado en --- "Clasificación automática et perception par Ordinateur", IRIA, pgs. 27 a 40, 1975.

