

INSPECCIONES POR ULTRASONIDOS ORIENTADAS POR LOS DATOS PARA MATERIALES COMPUESTOS

Juan I. Caballero¹, Carlos D. Martínez¹, Francisco Ballesteros²

¹Dep. Ciencia de materiales ETSI Caminos, canales y puertos. UPM. C/ Prof. Aranguren 3, 28040. Madrid.

Juanignacio.caballero@upm.es

²Dep. Matemática aplicada a las tecnologías de la información y las comunicaciones ETSIT. UPM. Av. Complutense, 30 28040 Madrid.

Resumen: En la fabricación de materiales compuestos se utilizan diferentes técnicas, principalmente no destructivas (NDT), para asegurar su calidad y ausencia de defectos tales como delaminaciones o presencia de porosidad. En los entornos de producción las técnicas preferidas son las basadas en ultrasonidos. Sin embargo, los ultrasonidos tienen limitaciones en la caracterización de los defectos presentes en un componente. En cambio, otra técnica: la tomografía computerizada por rayos-X (XCT) permite la reconstrucción de un componente y su interior con muy alto nivel de detalle. La adopción de la XCT en la industria se ve limitada porque se necesita acceder a las diferentes vistas de un componente, el elevado coste del equipamiento y la disponibilidad de profesionales con la especialización adecuada. Un enfoque reciente consiste en el uso de técnicas de ciencia de datos para mejorar la información de los ultrasonidos. En este trabajo de fin de máster se plantea una primera metodología con vistas a desarrollar técnicas de ultrasonidos orientadas por datos (*data-driven*). En particular, se propone enfocar la relación entre atenuación de los ultrasonidos y la presencia de porosidad como un problema de predicción supervisada de regresión para una pieza de CFRP ensayada a 1, 2.25, 5 y 10 MHz. Los datos de los ultrasonidos se emplean como variables regresoras y la información de la XCT, el nivel de porosidad, se toma como verdad de referencia (*ground-truth*). Los resultados muestran la viabilidad de la metodología, aunque trabajos futuros deben refinar los modelos y extender las variables para predecir otras características de los defectos.

Palabras clave: NDT, XCT, ultrasonidos, Machine learning, porosidad.

1. MOTIVACIÓN.

Los materiales compuestos precisan de técnicas que aseguren unos estándares de calidad. La porosidad es uno de los parámetros más sensibles ya que altos niveles pueden impactar en el desempeño en servicio de la pieza. Existen varias técnicas no destructivas (NDT) para la medición de la porosidad y otros defectos de las cuales destacan dos para este trabajo:

- NDT ultrasonidos: es el tipo preferido por su sencillez, coste y funcionalidad en servicio. Consisten en evaluar la señal de un ultrasonido en su propagación a lo largo del espesor de un componente. Si bien son eficaces detectando porosidad, no lo son tanto en su caracterización. Ver [1,2] donde comprueban que factores como tamaño y morfología de los poros influyen en el nivel estimado de porosidad.
- La tomografía computerizada por rayos X (XCT): Permite la reconstrucción en 3D con alta resolución. Consiste en la obtención de un número de radiografías realizadas a diferentes ángulos del componente y a las que se aplica un algoritmo para la generación del volumen.

Por un lado, las NDT por ultrasonidos otorgan sencillez y rapidez en la inspección *in situ*. Por otro, la XCT destaca en el alto detalle de la reconstrucción 3D. Sería deseable disponer de una tecnología con las ventajas

presentes en los ultrasonidos aportara el detalle de una XCT. La metodología que se expone propone cómo tratar los datos de las técnicas para entrenar un modelo de machine learning que permita predecir el nivel de porosidad y que sea ampliable a otros parámetros como el tamaño y distribución de los poros. La implementación consistiría en utilizar el modelo entrenado para predecir los detalles de la porosidad en componentes del mismo material a partir de su inspección de ultrasonidos, prescindiendo de la XCT.

El trabajo se plantea como un primer paso en desarrollar inspecciones por ultrasonidos mejoradas por machine learning. Se quiere exponer también el gran potencial que presentan las técnicas del machine learning para la ciencia e ingeniería de materiales, las técnicas NDT y su impacto en la sociedad por su relación con la industria 4.0.

2. INSPECCIÓN DE DEFECTOS.

En [3] se puede leer una review de los métodos NDE para la caracterización de la porosidad. Fue en los 70 cuando se estableció una relación bilineal entre la fracción volumétrica de poros y el coeficiente de atenuación de una onda de ultrasonidos. Se propuso un umbral en torno al 1.5% en torno al cual la atenuación cambiaba, en base a la evolución de los poros esféricos a elipsoides. Esos estudios permitieron el desarrollo de los equipos de inspección por ultrasonidos en materiales compuestos.

Desde entonces se han propuesto diferentes modelos según la morfología de los poros fundamentados por simulaciones en elementos finitos. Se ha establecido un umbral máximo del 2% de porosidad, por encima del cual el componente no se considera útil. Sin embargo, no se dispone de un “medidor universal” de la porosidad para cualquier material y las estimaciones varían en intervalos de incluso un 1% en función del tamaño y geometría del poro. [1,2]

Estudios recientes en otros tipos de defectos en compuestos han tomado vías de interés, diferentes y complementarias. En [4], Smith et al abordan la detección de pliegues o arrugas en las láminas de material compuesto. Para ello utilizan los datos del volumen 3D de los ultrasonidos. Además, no utilizan las representaciones de A, B o C-Scan, sino la señal analítica original (El A-Scan corresponde con la amplitud instantánea de la señal analítica). Obteniendo resultados para distinguir entre láminas de compuesto en base a la fase y frecuencia instantánea de la señal. Otro estudio, [5] Barry et al tratan también los ultrasonidos como un volumen. Fabricaron probetas con defectos introducidos artificialmente y construyeron un clasificador a partir de los volúmenes capaz de diferenciar entre los diferentes casos. Probablemente el futuro de las inspecciones por ultrasonidos evolucione hacia el uso del volumen en 3D, explotando atributos de la señal, además de la amplitud, y la integración de técnicas de machine learning para mejorar la detección y/o caracterización ya sea en regresión o clasificación.

3. DATA MINING AND DATA SCIENCE.

Debido a las limitaciones de espacio solo se realiza un resumen de los conceptos más importantes.

Ocurre frecuentemente en los campos más activos del conocimiento que es difícil establecer una definición concreta y este caso no es una excepción. El Knowledge Discovery in Databases (KDD) introducido en 1989 hace referencia al proceso global de identificar, validar y comprender patrones en los datos. Es decir, la transformación de datos *raw* en otras representaciones más explicativas o útiles. En esta tarea se emplean técnicas y nociones de la estadística, las bases de datos y/o el aprendizaje automático y profundo que se aplican dentro de una secuencia de pasos. El Data mining se referiría a uno de estos pasos, aunque recientemente ambos términos se usan de forma equivalente. El KDD se organiza en fases en un proceso iterativo no lineal en que existen dos estándares: El SEMMA y el europeo Cross-Industry Standard for Data Mining (CRISP-DM).

CRISP-DM

A continuación, se presenta cada fase que lo compone:

- *Business Understanding*: Consiste en establecer un objetivo, la motivación y los medios para alcanzarlo aunando aspectos funcionales y técnicos.
- *Data Understanding*: Trata la adquisición y análisis inicial de los datos.
- *Data preparation*: La principal fase en cuanto a volumen de esfuerzo, consume un 80% del tiempo

del proyecto. Consiste en el procesado y preparación de los datos para la fase de modeling.

- *Modeling*: En esta fase se aplican los modelos, es frecuente el uso de varios algoritmos y mecanismos de prueba y error en la optimización de sus hiperparámetros.
- *Evaluation*: Es necesario revisar los modelos desarrollados en nuevos datos.
- *Deployment*: Una vez que se han conseguido resultados, el conocimiento debe ser organizado y presentado de forma sencilla. Dependiendo del contexto puede materializarse en un informe, o en una aplicación disponible para toda la organización.

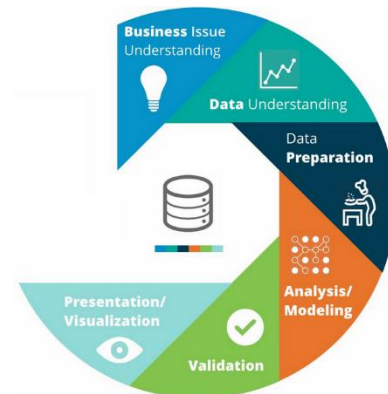


Figura 1: Esquema de fases del CRISP-DM.

Tipos de problemas de Data Mining:

Según el propósito se dividen en 2 bloques:

- Problemas descriptivos: De naturaleza más exploradora, consiste en encontrar patrones que describan los datos o ayuden a generar conocimiento.
- Problemas predictivos: El objetivo es construir un modelo que sirva en el futuro para predecir información.

Según la información disponible existen dos tipos:

- Aprendizaje supervisado: Consiste en entrenar modelos en base a que generen resultados lo más próximos posible a unos valores objetivo conocidos de antemano. A su vez consta de dos tipos:
 - Estimación: La variable a predecir es continua.
 - Clasificación: La variable objetivo es discreta, categórica.
- Aprendizaje no-supervisado: Consiste en encontrar patrones en datos donde no se dispone de valores objetivo.

Las características y su selección.

La selección de las variables para predecir en un es crítico para el éxito de los proyectos de *Data Mining*. La selección de características es un proceso donde se reduce la cantidad de datos a analizar, centrando el análisis en los datos importantes y mejorando la calidad del conjunto. Existen familias de técnicas para ejecutar la selección de características que por cuestiones de espacio no se incluyen.

Modelos en estimación y aprendizaje supervisado:

Existe consenso en que no existe un modelo que aproxime mejor en todos los casos y la selección se basa en un mecanismo de prueba y error. Por motivos de espacio solo se menciona el modelo que mejor funcionó:

- Los árboles de regresión son técnicas en los que se jerarquizan las características en base a cuánto disminuyen una determinada métrica de información existen varias opciones que escapan al alcance de este artículo.

Validación

Es crucial una vez desarrollado un modelo determinar si este es válido en un conjunto de datos independientes del data set original.

Existe el concepto del balance sesgo-varianza. El sesgo es el error cuando se aproxima un modelo complejo con uno simple. La varianza se define como el error del modelo entre diferentes conjuntos de datos. Así, pueden ocurrir predicciones con *underfitting* donde el modelo usado es demasiado sencillo para el problema a resolver, y otras veces puede ocurrir *overfitting*, donde el modelo se ajusta a un determinado conjunto de datos y generaliza mal para nuevos conjuntos.

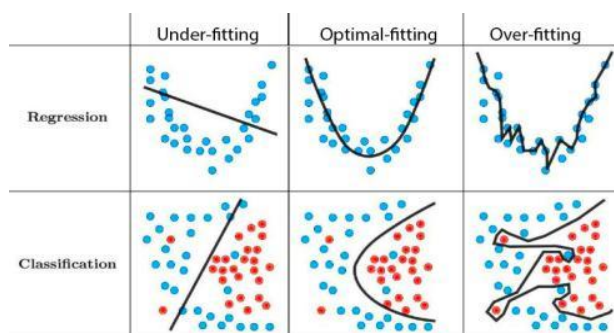


Figura 2: Ejemplos de sesgo-varianza

4.- MATERIALES Y DATOS.

Se obtuvo una probeta de 40x80x5 mm² formado por 16 láminas de prepreg M56 en diferentes orientaciones y curadas con bolsa de vacío.

La probeta fue inspeccionada a 1, 2.25, 5 and 10 MHz y se obtuvieron representaciones C-Scan de 375x101.

Mediante la XCT se reconstruyó el volumen 3D tras la concatenación de tres medidas. El volumen final cuenta con unas dimensiones de 2040x4020x280 y 16 bits de grises.

5.- SOLUCIÓN PROPUESTA.

El objetivo del trabajo consistía en establecer una metodología orientada por las fases de CRISP-DM que permitiera abordar la relación entre la porosidad en materiales compuestos y la atenuación de la onda de ultrasonidos desde el prisma del KDD. En los siguientes puntos se realiza un resumen de lo más importante siguiendo las fases de KDD.

Business Understanding

Con el fin de entender la fase de *data preparation* conviene detallar cómo sería el data set al final de la fase. En disciplinas donde el KDD está más extendido, como la medicina, un dataset es parecido al siguiente ejemplo:

Tabla 1: Data set de ejemplo

	Sexo	Peso (Kg)	Talla (cm)	...	Enfermedad S/N
Paciente 1	V	60	180	..	S
Paciente 2	M	100	165	..	S
Paciente 3	V	80	172	..	N
Paciente 4	V	100	200	..	S
Paciente 5	M	50	155	..	N
Paciente 6	V	70	160	..	N
Paciente 7	M	65	165	..	S
Paciente 8	M	90	158	..	N

Se encuentran tabuladas una serie de características en las columnas (atributos) y los valores en cada fila (registro). Los atributos en color azul son las características regresoras. La naranja, la variable objetivo o el *target*. Cada registro corresponde a los valores de cada atributo para un paciente. Basado en esta idea se planteó un data set en el que la variable *target* sería el nivel de la porosidad y las variables regresoras características extraídas de los ultrasonidos. Los atributos más directos son los valores de atenuación de cada C-Scan y el nivel de porosidad correspondiente en la XCT como muestra la tabla 2.

Tabla 2: Data set esperado

	Valores de Atenuación (db)		Var5	Var4	Porosidad (%)
ID_1	15	17..			0.5
ID_2	12				1.24
ID_3	9				0.7
ID_4	6				0.75
ID_5	16				0.25
ID_6	14				0.32
ID_7	11				0.15
ID_8	9				0.89

Ultrasonidos
XCT

Cada registro representa una fracción de la probeta, que por sencillez se decidió a la columna de base 1mm² equivalente un pixel de C-Scan. En conclusión, igual que en el data set médico cada registro corresponde con un paciente y sus diferentes atributos de salud, en el data set de materiales cada registro corresponde a una fracción del volumen y sus atributos de las inspecciones. Por sencillez, la fracción de volumen tiene la base equivalente a un pixel del C-Scan y la altura es el espesor del laminado.

Data preparation

Por cuestiones de espacio solo indicar que el volumen de XCT se segmenta para destacar únicamente los poros. Los ultrasonidos tienen que recortarse para establecer una relación entre los píxeles y vóxeles de todas las imágenes.

Se desarrolló un script para calcular los diferentes atributos. Los cuales se envían a una base de datos.

Modeling

Lo primero consiste en la formación del data set llamando a la base de datos y realizando una selección de las características.

Los modelos elegidos para probar fueron la regresión lineal, los árboles de regresión y un random forest. El

motivo era elegir modelos que permitan la interpretación de los resultados.

Validation

Se escogió una técnica denominada K-fold cross validation para entrenar el modelo. Sin embargo, se percibió que los registros con baja porosidad son mucho más frecuentes que los de alta, que son los de mayor interés. Existen métodos para tratar los datos cuando las variables objetivo presenta estas diferencias que no se han podido detallar en este artículo.

6.- EVALUACIÓN DE LOS RESULTADOS

Las predicciones no superan a las de los modelos del estado del arte. Se estima el nivel de porosidad en un intervalo con mínimo cercano al 0.2 % y un máximo de 0.8 % en el agregado de los registros.

El error es mucho mayor si se entra en regiones locales de la probeta. La ilustración 3 es una gráfica habitual en la ciencia de datos. En ella se compara la variable objetivo (el nivel de porosidad) medida en el eje x, con el valor obtenido por el modelo entrenado en el y.

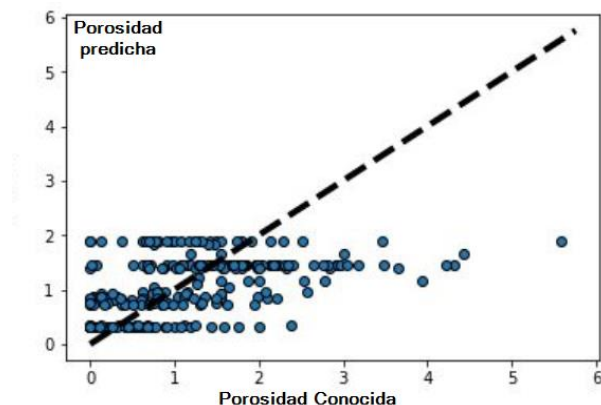


Figura 3: Comparación entre la porosidad real y la estimada por el modelo.

En este caso cada punto representa una de las fracciones del volumen (un registro). Idealmente los puntos deberían situarse en las inmediaciones de la recta identidad. Por el contrario, se observa cómo hay registros con un nivel de porosidad real de 0 % a los que el modelo predice el 1% o incluso el 2%. Asimismo, el modelo no estima por encima del 2%.

Se observó que los ultrasonidos no aportan el contraste entre píxeles vecinos que permite la XCT. Se probó a cambiar el tamaño de la fracción volumétrica de los registros. En la ilustración 4 los registros equivalen a una base de 3mm². De esta forma ajusta mejor en el intervalo de 0-1.5% de porosidad, aunque sigue sin aprender los registros de mayor porosidad.

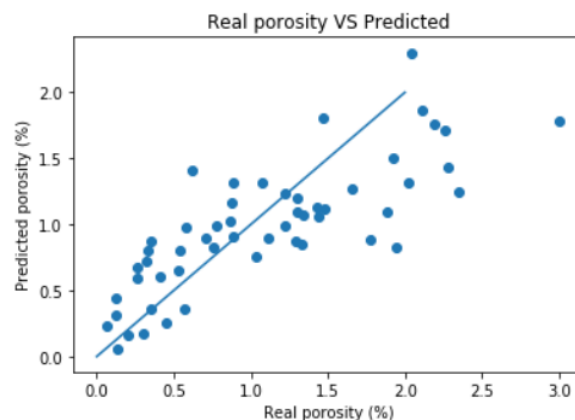


Figura 4: Comparación entre porosidad medida y estimada

7.- CONCLUSIONES.

Se presenta un caso de aplicación del KDD para la ciencia de materiales: la medición del nivel de porosidad en materiales compuestos. Se propone una metodología que combina dos técnicas de inspección complementarias: los ultrasonidos y la XCT siguiendo las fases del estándar en la disciplina, el CRISP-DM.

Los resultados muestran un error global cercano a los del estado del arte, sin embargo, el error local es muy alto para ciertos registros. No obstante, la metodología abre líneas de trabajo a futuro como ampliar el volumen de datos, las características y extender la predicción a otras variables de la porosidad.

Para más detalle se puede solicitar el trabajo de fin de máster. El trabajo a futuro se desarrollará como parte del doctorado de Juan I. Caballero.

8.- AGRADECIMIENTOS.

Agradecer la colaboración en el proyecto de Ernestina Menasalvas y Consuelo Gonzalo del departamento MIDAS de la UPM y de Federico Sket del IMDEA Materiales.

9.- REFERENCIAS.

- [1] Shanshan Ding et al. "Investigations on Relationship between Porosity and Ultrasonic Attenuation Coefficient in CFRP Laminates Based on RMVM" In: *International Symposium on NDT in Aerospace*,
- [2] Li Lin, Ming Luo, and Hongtao Tian. "Experimental Investigation on Porosity of Carbon Fiber-Reinforced Composite Using Ultrasonic Attenuation Coefficient" In: *World Conference on Non destructive Testing, Shanghai, China*.
- [3] E A Birt and R.A Smith "A Review of NDE Methods for Porosity". In: *Insight – Non-Destructive Testing and Condition Monitoring* 46.11
- [4] Robert A. Smith et al. "Ultrasonic Analytic-Signal Responses From PolymerMatrix Composite Laminates". en. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 65.2 (Feb. 2018)
- [5] Tj Barry et al "Defect Characterisation in Laminar Composite Structures Using Ultrasonic Techniques and Artificial Neural Networks" In: *Journal of Composite Materials* 50.7