

# Configuración automática de una máquina de fisioterapia para la espalda. Estimación de datos iniciales

González Domínguez, Diego

**Resumen.** Para la configuración de la máquina de fisioterapia, se utilizarán datos de pacientes anteriores, encontrando el/los caso/s de mayor similitud para cada nuevo paciente, y obteniendo los parámetros que mejor se ajusten a las características del problema en cuestión. Sin embargo, debido a problemas en las mediciones, no se dispone de datos completos para realizar dicha configuración, por lo que el primer paso para llevar a cabo dicho proyecto será hacer una estimación de algunos de los resultados experimentales.

## 1 - INTRODUCCIÓN

La máquina de fisioterapia sobre la que trata este artículo consiste en una serie de servos, arneses y sensores que permiten al paciente ejercitar las zonas doloridas. Al mismo tiempo, el personal médico puede medir el grado de mejora de los pacientes a partir de unos test de dolor y la fuerza ejercida por éstos sobre los mecanismos de la máquina.

Para llevar a cabo correctamente el procedimiento, la máquina debe configurarse previamente de acuerdo a las características de los pacientes y sus dolencias, lo cual requiere bastante tiempo por parte del personal médico. Para facilitar y acelerar el proceso, se realizará una configuración automática a partir de una base de datos disponible mediante los resultados obtenidos con pacientes anteriores.

Sin embargo, debido a errores en las mediciones, fallos al guardar los datos de la aplicación, o a que algunos pacientes no han podido asistir a todas las sesiones de fisioterapia correspondientes, los datos experimentales presentan huecos que pueden ocasionar errores importantes en la configuración automática, por lo que el primer paso a seguir será hacer una estimación de dichos parámetros. Para ello se dispone de los datos de los pacientes (sexo, edad, parámetros físicos, etc.), y se pretende estimar los resultados de evaluación del dolor (EVA) y fuerza ejercida sobre la máquina (presiones) que no hayan sido tomados correctamente.

Para realizar dicha estimación, se utilizarán los casos de mayor similitud, los cuales se determinarán a partir de la semejanza de los resultados de otros pacientes con el caso actual y ordenándolos en función de ésta. Tanto para la estimación como para la ordenación se emplearán diferentes técnicas, por lo que será necesario realizar una comparación entre los diferentes métodos, eligiendo el que produzca mejores resultados.

En ambos casos (resultados de EVAs y presiones) el procedimiento es análogo, pero con unas pequeñas variaciones debido a las diferencias en la toma de datos. Se estudiará cómo obtener los datos de EVA, y se añadirá la modificación que se debe hacer para aplicar el mismo proceso a los datos de presiones.

## 2 – MÉTODOS DE ESTIMACIÓN DE DATOS

El problema a resolver consiste en que se dispone de un vector numérico con huecos que se deben rellenar. En primer lugar, se debe elegir un método para realizar la estimación:

**a) Extrapolación sobre la serie:** La extrapolación consiste en considerar que los datos siguen un patrón, y que dicho patrón se extiende más allá de los datos conocidos. El método trata de encontrar la función que mejor ajuste los resultados disponibles y la aplica a los huecos.

Existen distintos métodos de extrapolación a considerar, dependiendo del tipo de función con la que se realice el ajuste:

- Lineal: método regresión lineal
- Polinómica: método de Newton de diferencias finitas.
- Exponencial: método de extrapolación de Curva Francesa

Este método no usa conocimiento adicional, por lo que no requeriría analizar el resto de datos experimentales y en consecuencia sería relativamente rápido. Sin embargo, debido a la variabilidad de los datos, esto también podría conllevar un mayor error en la estimación, ya que es difícil tener en cuenta tendencias, ciclos, patrones o irregularidades. En cualquier caso, se implementará este método para poder compararlo con los demás.

**b) CBR (Case-Based Reasoning) simple:** El método de CBR, o razonamiento basado en casos, consiste en resolver un problema a partir de otros ya resueltos. En el caso que nos ocupa, consistiría en estimar las series de datos incompletos de un paciente a partir de los resultados completos de otros pacientes.

En el CBR simple los datos a considerar serían solo las series de resultados experimentales (que serían, por tanto, a la vez descripción y solución). Para encontrar el/los paciente/s cuyas características sean lo más parecidas posibles al caso de estudio (y, por tanto, tendrán una mayor probabilidad de tener una salida similar), se comparará el paciente actual  $i$  con otro  $j$  con resultados conocidos mediante una función de similitud de la forma

$$F(i, j) = \sum_{k=1}^N x_k \cdot f_k(i, j) \in [0,1] \quad (1)$$

donde hay  $N$  datos experimentales conocidos, los  $x_k$  son los pesos relativos del dato  $k$  en la formula, tales que

$$\sum_{k=1}^N x_k = 1 \quad x_k \in [0,1] \quad (2)$$

y los  $f_k(i,j)$  son funciones de similitud entre los pacientes  $i$  y  $j$  para el dato  $k$ , con  $f_k(i, j) \in [0,1]$

El valor de cada  $x_k$  dependerá de la importancia que le quiera dar a cada parámetro. En el caso de las series de datos, se asume que todos tienen la misma importancia, por lo que  $x_k = 1 / N$ .

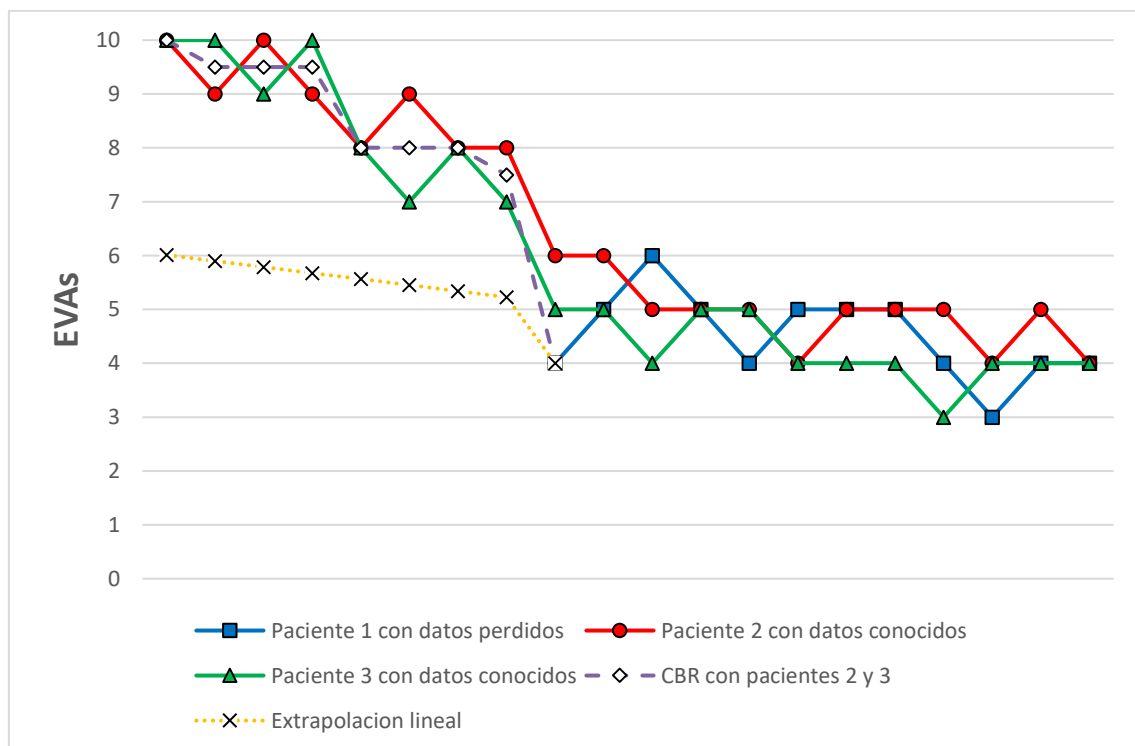
Así mismo, la función de similitud entre datos continuos (como los de los resultados experimentales con los que se trabaja) puede basarse en el porcentaje de variación entre ambos:

$$f_k(i, j) = 1 - \frac{|k_i - k_j|}{\max(k_i, k_j)} \quad (3)$$

siendo  $k_i$  el valor del  $k$ -ésimo dato conocido del paciente  $i$ .

El objetivo es obtener una lista ordenada de pacientes  $a, b, \dots, z$  tales que  $F(i,a) \geq F(i,b) \geq \dots \geq F(i,z)$ . Con esta lista, se determinará que paciente/s será/n los de mayor similitud, pudiendo considerarse varias opciones:

- El de mayor función de similitud



### 3 – DATA SET

Los datos con los que se trabajará están en el archivo DATOS.xlsx, siendo los datos de entrada los correspondientes a las pestañas *Datos* y *SF36* (estadísticas físicas), mientras que los datos incompletos y que se busca estimar son los de las pestañas *EVAs* (evaluación del dolor) y *Presiones*.

En ambos casos el procedimiento es análogo, pero con unas pequeñas variaciones debido a las diferencias en la toma de datos: en *EVAs*, los huecos se encuentran al principio de las series de datos, mientras que en *Presiones* están al final.

Un problema que podemos encontrar es que tampoco todos los datos de entrada están disponibles en algunos casos, por lo que habría que hacer una estimación previa, o bien considerar no tenerlos en cuenta para la estimación de los EVAs y presiones.

Además, algunos datos *SF36* también están incompleto, por lo que, si realizase el mismo proceso para completarlos, no sería válidos para incluirlos en la estimación de los datos experimentales. Como se dispone de 5 mediciones, podría considerarse hacer una media de todas ellas (o de las que estén disponibles en cada caso), de manera que todos los pacientes dispusiesen de datos completos. De manera similar, existen parámetros en *Datos* que deberían ser completados para poder ser empleados en algunos métodos de estimación.

Por último, un problema importante es que inicialmente solo se dispone de muy pocos pacientes con todos los datos EVAs completos, por lo que, si su similitud con los demás pacientes no es muy alta, la estimación de los parámetros de salida no será muy precisa. Una posible solución para ello sería ir completando los datos de los pacientes en orden creciente según el número de datos sin completar, de manera que cuantos más datos se requieran haya más casos completos disponibles.

En cualquier caso, para poder trabajar con los datos, se exportará cada pestaña a un archivo con el mismo nombre y la extensión .csv. Además, se debe hacer una pequeña corrección en los datos de SF36.csv, ya que uno de los datos aparece como un espacio en lugar de estar vacío, lo cual podría ocasionar problemas en las siguientes etapas.

### 4 – EVALUACIÓN DE LOS MÉTODOS DE ESTIMACIÓN

Con los diferentes métodos que se han sugerido tanto para la función de similitud como para la estimación de los resultados experimentales, es necesario evaluar la optimalidad de cada caso. Como en el apartado anterior, se pueden considerar varios métodos para realizar la evaluación:

**a) Evaluación de series incompletas:** Esta evaluación consiste en tomar una serie de datos conocidos, eliminar una cantidad aleatoria de datos ( $K$ ) y rellenarlos usando alguno de los métodos de estimación propuestos. A continuación, se compararía el resultado con la cadena original y se calcularía su grado de similitud (de manera similar al método del CBR simple).

Repetiendo el proceso para diferentes cadenas de datos experimentales y valores de  $K$ , se caracterizarán los distintos métodos de estimación en función del porcentaje de huecos que admiten las series de datos, es decir, el  $K$  máximo a partir del cual el grado de similitud con la cadena original no supera un determinado umbral (fijado previamente).

**b) Evaluación con datos del paciente incompletos:** Para esta evaluación se procede de manera similar al método anterior, pero en este caso se usarán datos de pacientes incompletos (por fallos del lector, errores al guardar, etc.), y que de otra forma no se podrían utilizar. Esto no afectaría a los métodos de estimación por extrapolación y CBR simple, pues no tienen en cuenta los parámetros de los pacientes, pero si puede influir en el de CBR completo a la hora de obtener la función de similitud ampliada.

Las opciones a considerar son:

- No tener en cuenta los datos desconocidos (usar  $x_k = 0$  para el  $k$ -ésimo parámetro si este no se conoce en alguno de los dos pacientes que se comparan)
- Estimar el valor ausente. Como se realizan diferentes mediciones de los distintos parámetros, los datos que faltan pueden estimarse mediante la media de otros valores que sí se conocen, por regresión, etc. No sería correcto usar el método CBR en este caso, puesto que entonces incluir los datos estimados de esta manera en la estimación posterior de los datos experimentales podría provocar un mayor error.

Aplicando el método de evaluación cruzada, se determinarán los métodos más eficientes de entre todos los propuestos: se repite el proceso anterior para todas las combinaciones posibles entre funciones de similitud y de evaluación, y con diferentes cadenas iniciales. Además, como se comentó anteriormente, se calcularán los coeficientes de la función de similitud mediante un algoritmo genético.

Para terminar, habiendo determinado tanto el método de estimación como los coeficientes de la función de similitud, se emplearán estos para calcular los huecos en los datos experimentales, y poder continuar con el proyecto de calibrado de la máquina de fisioterapia.