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

González Domínguez, Diego

**4 – EVALUACIÓN DE LOS MÉTODOS DE ESTIMACIÓN**

El objetivo de esta evaluación es encontrar el método que mejor ajuste los datos experimentales disponibles y utilizarlo para estimar los datos de aquellos pacientes cuyas mediciones no están completas.

Para ello, se han implementado los diferentes métodos de estimación y se han evaluado según el siguiente criterio:

1. Se toman los datos de un paciente
2. Se elimina una parte de los datos de ese paciente
3. A partir del método seleccionado, se estiman los datos eliminados.
4. Se evalúa la similitud de los datos estimados respecto a los reales.
5. Se repite el proceso un número aleatorio de veces para el mismo paciente y eliminando diferentes cantidades de datos.
6. Se repite para todos los pacientes.

Realizando la media para todas las evaluaciones de una paciente y de todos los pacientes, se han obtenido una serie de resultados que se detallan a continuación.

(Los resultados se corresponden con el análisis de las Presiones, pero son análogos a los de *SF36* y *EVA*s. Además, no se han tenido en cuenta a algunos de los pacientes ya que se dispone de muy pocos datos de ellos)

* **Métodos de extrapolación:**

Estos métodos solo tienen en cuenta los datos a estimar, y mediante funciones de regresión se obtienen los datos que faltan. Comparando estas estimaciones con los datos iniciales, se obtienen os siguientes resultados:

A la vista de los resultados, el método más eficaz para estimar los datos de los pacientes parece ser el método lineal (similitud media del 93,5%) o el exponencial (93,7%), mientras que el método polinómico es ligeramente peor (91,7%). Sin embargo, el método exponencial no resulta aplicable a los casos de *SF36* y *EVAs*, ya que en estos casos hay datos negativos que no pueden obtenerse por este método, por lo que el método más eficaz entre éstos será el lineal.

* **CBR simple:**

Para estos métodos, se han tenido en cuenta solo las cadenas de datos a estimar (Presiones, SF36 y EVAs), sin tener en cuenta el resto de parámetros físicos de los pacientes. En cada caso, se han tomado los datos del paciente más próximo, o de los dos, tres o cinco más próximos:

En todos los casos se obtiene una similitud media similar (95,9%, 95,5%, 95,4% y 95,2%, respectivamente), pero es evidente que cuantos más vecinos se usen para estimar los datos de un paciente, mayor será la desviación, por lo que, a priori, el mejor método de CBR simple será el que utiliza solo el paciente con mayor similitud al caso estudiado para estimar los datos.

* **CBR completo:**

Mientras que en el CBR simple solo se tenían en cuenta los datos a estimar para hacer la comparación entre pacientes, en el CBR completo también se tienen en cuenta los datos físicos y otras características de los pacientes.

Sin embargo, algunos de estos datos también están incompletos y deberían rellenarse antes, pero si se utilizase también aquí el método CBR afectaría a la estimación posterior, por lo que se han utilizado dos métodos diferentes: extrapolación lineal y valores medios.

Con la extrapolación lineal se realiza una regresión lineal de los datos disponibles y se calculan los que faltan a partir de ella:

En este caso, las similitudes medias para los métodos con uno, dos, tres o cinco vecinos son 94,8%, 94,7%, 94,9% y 94,9%, respectivamente, lo que quiere decir que, aunque los resultados son muy similares, en este caso se invierte la tendencia del CBR simple y cuantos más vecinos se utilicen para hacer la estimación, mejor.

El otro método que se ha utilizado, ha sido tomar los valores de cada medición, hacer la media, y rellenar tanto los valores disponibles como los que faltan con esa media:

De nuevo, los resultados son muy similares (similitudes medias de 94,6%, 94,7%, 94,7% y 94,7%), si bien cuanto más vecinos se utilicen, parece que también es mejor la estimación.

Finalmente, comparando los métodos con mejores resultados en cada apartado, se tiene:

En vista de los resultados, se deduce que el mejor método de estimación es el de CBR simple, con un vecino, con un 95,9% de similitud media.

Es importante decir que para los métodos de CBR completo se han utilizado unos coeficientes para ponderar la importancia de cada uno de los datos en la función de similitud. En principio, se pretendía utilizar un algoritmo genético para optimizar dichos coeficientes y mejorar el método CBR completo, pero en la práctica no se ha observado una mejora substancial (en torno al 0,8% de mejora en la similitud, pero aun así por debajo debla obtenida con el CBR simple) e implicaba una maximización del coeficiente correspondiente a las Presiones (o SF36 o EVAs en cada caso), lo que lo equivaldría a que la máxima similitud del método CBR simple se obtiene cuando más se asemeja al CBR simple.

Aquí puede verse una evolución del algoritmo genético utilizado y como la mejora obtenida es mínima respecto al tiempo utilizado (unas 10 horas por ejecución):

Por tanto, la conclusión a la que se llega es que el mejor método para estimar los datos de Presiones que faltan para ciertos pacientes es el de CBR simple, y utilizando solo los datos del paciente con mayor similitud. Se llega a la misma conclusión para SF36 y EVAS, aunque con una similitud menor que en el primer caso (92,6% y 93,2%, respectivamente).