La interpolación y el clima

Departamento de ingeniería de sistemas. Pontifia universidad Javeriana.

Andrés Mauricio García Mutis - Estudiante ingeniería de sistemas - ga_andres@javeriana.edu.co

José David Calderón Benavides - Estudiante ingeniería de sistemas - j_calderon@javeriana.edu.co

Luis Felipe Ariza Ortiz - Estudiante ingeniería de sistemas - arizaluis@javeriana.edu.co

William Orlando Rodriguez - Estudiante ingeniería de sistemas - rodriguez_william@javeriana.edu.co

***Resumen*— Mediante el uso del método numérico conocido como interpolación y spline cúbico para predecir la temperatura interior dados el día y la hora del año 2013, nuestro modelo predice la variable de temperatura con base en datos de una región cercana en el área de Fortaleza (Brasil).**

***Palabras clave*- interpolación, spline, método numérico, estación meteorológica.**

I. Introducción

La predicción y el análisis de datos es aplicado para el desarrollo de modelos climáticos usados en los institutos meteorológicos, ya que basándose en las lecturas de variables climáticas en diferentes localizaciones geográficas que presenten condiciones similares en el terreno se pueden aproximar dichas variables basándose en las lecturas obtenidas. El principal uso que se le dan es para complementar la información faltante de una zona en la que no se presente una estación meteorológica. Esto nos servirá para poder completar la información faltante de la temperatura interna de la región que seleccionamos que en este caso fue Jati, y el método que vamos a usar para esto es la interpolación por spline cubico.

II. Datos seleccionados

Para la selección de datos se contaba con las mediciones de 17 estaciones donde estas mediciones pueden variar entre 164 hasta 720 entre cada estación. Se tienen que elegir dos estaciones donde una seria representada como la fuente del modelo y la otra seria representada como el objetivo de la aplicación del modelo. Para que el modelo fuera lo más exacto posible, se necesitaba que los datos sean exactos y con el menor error posible para esto las estaciones debían estar cercas unas a otras, por eso se seleccionó a crato y jati ya que estaban cerca.

Respecto a las variables que se tienen en las estaciones, se eligió la temperatura interior ya que se tenían variables completas sin tener valores nulos.

Por último, los datos que tratamos de la estación de crato se omitieron aquellos que tenían un valor inferior a -50 grados centígrados, ya que esta presento problemas a la hora de ser usado en el modelo.

A. *Temperatura según el tiempo*

Para realizar un análisis de la temperatura, elegimos algunos de los pares ordenados sacados de los resultados del filtro de datos y la construcción de los archivos. A continuación, se presentarán graficas donde se observan los respectivos comportamientos relacionados a las estaciones relacionadas. La grafica siguiente representa la temperatura interna con respecto a el tiempo en Crato.

A continuación podemos observar la grafica arrojada por el algoritmo

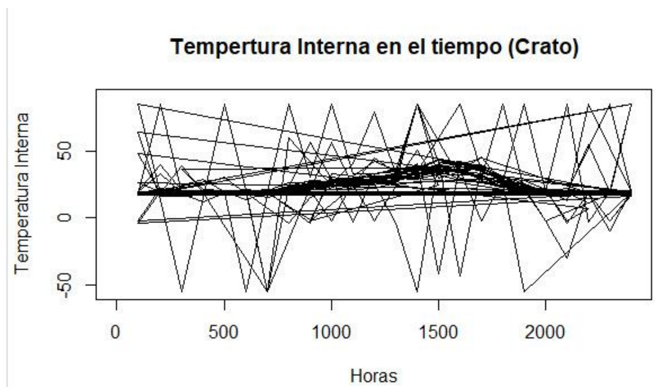

Ilustración 1 Temperatura interna en el tiempo para Crato

Podemos observar que en algunos de los puntos de la gráfica no se realiza o se hallan muestras dándole sentido a ciertos puntos de la gráfica.

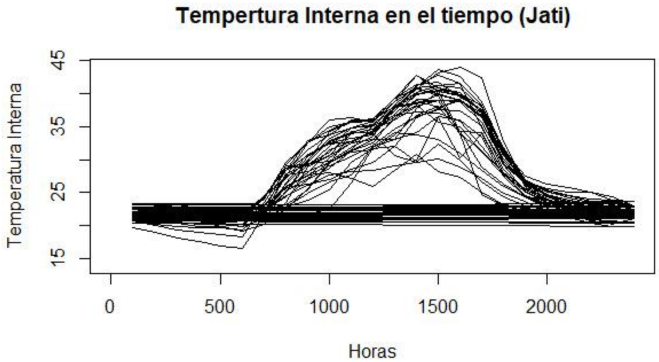

Ilustración 3 Temperatura interna en el tiempo para Jati

B. Interpolación

La conducta de los polinomios de nivel elevado oscila lo que se adapta al comportamiento de los datos que tomamos para llevar a cabo la interpolación por lo cual una interpolación polinómica es pertinente en esta situación.

Gracias a que los polinomios de grados bastante elevados tienden a oscilar, decidimos partir el intervalo atendiendo de esta forma a un modelo de aproximación polinomial por tramos para llevar a cabo el modelo que representara la conducta de la temperatura en la época.

La representación más sencilla de esta clase de interpolación podría ser la lineal debido a que esta va uniendo los puntos de vista daros con rectas, pue para producir las ecuaciones de una recta no es necesario más que 2 puntos de vista. Finalmente resulta un modelo con picos y curvas poco suaves, con un aspecto parecido a las gráficas de aspectos presentadas anteriormente. Empero la temperatura no se comporta de dicha manera continuamente va gradualmente, no existe tal cosa como un cambio bien sea incremental o decrementa de la temperatura ocurriendo paralelamente, por consiguiente, su funcionalidad no puede tener picos pues ello querría mencionar que hay 2 pendientes en un cierto punto.

Según la interpretación expuesta anteriormente la interpolación lineal no representa el comportamiento de la temperatura respecto a el tiempo. Por lo tanto, optamos por la utilización del spline cubico teniendo en cuenta que un polinomio general implica cuatro constantes para tener certeza de que el interpolante no puede ser diferenciado en el intervalo, así incluyendo una segunda derivada continua.

Para la utilización de los spline, se hizo uso de la librería splines implementada en R, la cual por medio de la utilización de nodos construye un spline cubico. De esta manera dando como resultado parejas ordenadas, siendo estas los datos de temperatura tomados en los instantes dados. Por parte de los polinomios utilizados, empleamos splinefun que elabora la función por medio de la utilización de los puntos. Cabe resaltar que utilizamos 469 nodos para hallar el spline por lo que hacer uso de funciones de la librería de splines que retornan matrices de valores no seria adecuado para nuestro propósito.

En la siguiente imagen podemos observar de manera clara la aproximación hallada por medio de los splines cúbicos a partir de los nodos.

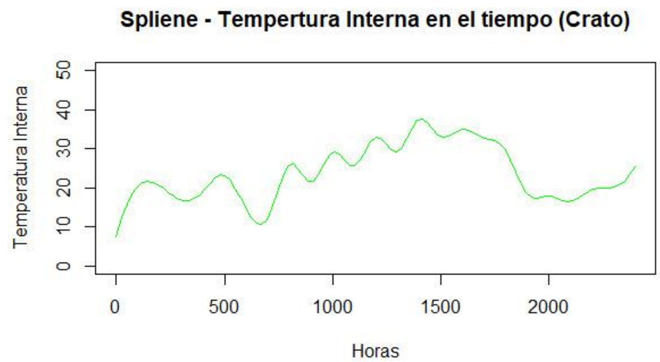

Ilustración 3 Splines cubicos hallados en la temperatura interna en el tiempo para Crato.

C. Modelo

Como pudimos observar en el spine cubico presentado anteriormente, al utilizar 469 nodos contrastamos en el comportamiento con un total del 65,2% sobre el total de putos.

ERROR

Para hallar el error entre nuestros propios datos vigilados y nuestra estimación polinómica, hicimos una comparación entre ambos aspectos por punto. Al ingresar un costo inicial a nuestra funcionalidad polinómica, en esta situación, las horas escogidas en nuestro modelo de tendencia (un 20% de los datos, escogidos aleatoriamente), tenemos la posibilidad de hallar la predicción de la temperatura con nuestro polinomio estimador.

Después, se presentará una gráfica con el error encontrado punto por punto, donde el costo Y es el error y el costo X son las horas en la que el error se calculó.

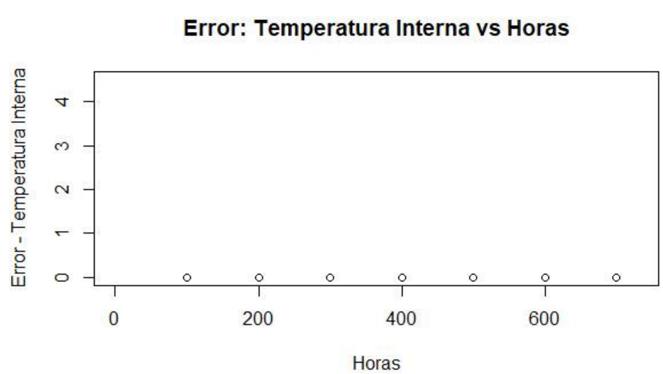

Ilustración 4 Error de la temperatura interna respecto a las horas

Como deseado, tenemos la posibilidad de ver ciertos valores atípicos presentes en la gráfica. Esto se debería a la naturaleza de los datos, debido a que dichos poseen sus propios datos atípicos. Si hacemos referencia a la Ilustración 4, tenemos la posibilidad de ver que los aspectos donde hay más error es en donde hay más espacio entre horas. Para hacer más claro esta comparación, hicimos una interpolación de nuestro error, utilizando el procedimiento de splines cúbicos:

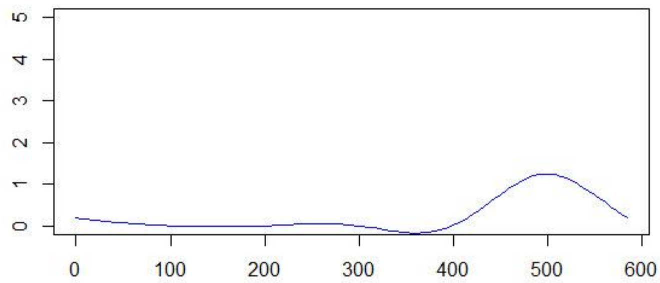

Ilustración 5 Interpolación de Error Punto por Punto.

Se puede ver que, cerca de las 200 horas y las 600 horas hay los valores máximos. Si la comparamos con la ilustración 5 de a partir de archivo, tenemos la posibilidad de ver que estas son las zonas en las que faltaban más datos para hacer la interpolación. Ahora, para hacer un estudio completo del error encontrado, hicimos 2 cálculos: el Error Cuadrático Medio y un Error Absoluto Medio:

```
A continuación, tenemos el Error Cuadrático Medio:
> MedioCuadrado = ErrorCuadraticoMedio(h)
> cat(MedioCuadrado)
0.3537633
> cat("\nError Absoluto Medio: ")

Error Absoluto Medio:
> cat(NoPromedio)
0.05498743
> errorPuntoVecino <- ErrorPuntoPorPuntoVecino(h)
```

Ilustración 6. Resultados para ECM y EAM.

Acá tenemos la posibilidad de ver 2 resultados diversos sobre nuestra medición de error. El error cuadrático medio resulto con alrededor de 0.69 más en su costo que en el error absoluto medio. Esto se debería, a que, generalmente, el error cuadrático medio les da más hincapié a los valores atípicos, o “outliers”, en el error. Esto quiere decir que generalmente, hay bastantes valore atípicos presente en nuestro error. Para concluir el estudio de errores, haremos un estudio con base en diversos tipos de errores:

- Error máximo: Está que el error máximo encontrado en la diferencia entre el modelo interpolado y las gráficas clásicos está cerca de los valores de 210 horas y la diferencia se marca con un costo de 13 para los datos interpolados y 50 para el costo de los datos habituales dando como consecuencia 10 de error en medio de éstos 2, aun cuando la diferencia podría ser además relacionada a que el pico de la temperatura no interpolada podría ser más grande unas escasas horas después del pico de la interpolada aun cuando la interacción es la misma.
- Error mínimo: Este error está cerca de los valores de 500 horas donde está la menor diferencia entre los valores de temperatura interpolada y no interpolada por la parte baja de las gráficas con valores de 13 en la gráfica interpolada y 22 en la no interpolada dando una diferencia de 1 siendo el error mínimo de entre los valores.
- Error medio: Si se suman los valores de error del mínimo (1) y el máximo (10) dividido entre 2 y daría como resultado 5.5
- Error absoluto: Seria el resultado de la suma de todos los valores de error dejando como resultado 13.00996 con todos los valores positivos.

VECINO MAS CERCANO

En algunas ocasiones al querer hacer la medición climita de un espacio se vuelve imposible, así sea ya que el sector es de difícil ingreso o la máquina de medición se descompuso, para dichos casos existe el procedimiento del vecino más cercano que es procedimiento para clasificar casos basándose en su parecido a otros. Se desarrolló como una forma de reconocer patrones de datos sin la necesidad de una coincidencia precisa con patrones o casos almacenados, donde los casos parecidos permanecen próximos y los que no lo son permanecen alejados entre sí.

En esta situación se hace el procedimiento en medio de las estaciones de Crato y Jato, donde se asumirá que los datos de la estación de Crato son igual al del spline de la estación Jato, dando como consecuencia la siguiente gráfica:

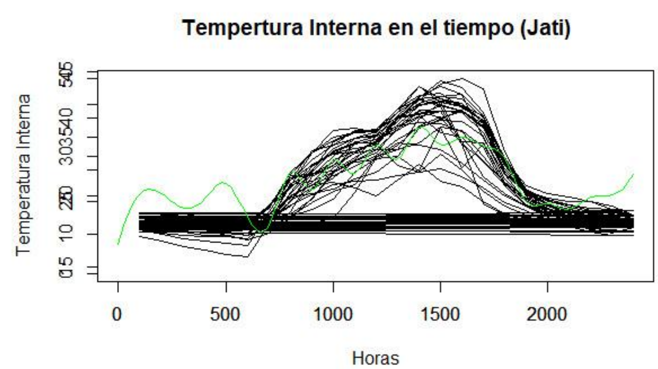

Ilustración 7 Temperatura interna en el tiempo para Crato y spline de Jato

De la gráfica tenemos la posibilidad de mirar que en las horas donde se poseen datos de la estación de Crato el spline de la estación de jati acierta en los incrementos y descensos de la temperatura, no obstante, en ciertos aspectos máximos y mínimos se equivoca, dando un margen de error. Sin embargo, donde más se diferencian las gráficas son en las horas donde no se poseen datos de la estación de Crato, para esto sirve el procedimiento del vecino más cercano, puede decirse que los valores del spline en aquellas horas son cercanos a los valores reales, precisamente esto todavía no se puede aceptar debido a que no se han calculado los errores.

Error

Para calcular los errores solo se tomaron en cuenta las horas donde se tienen los datos de la estación de Crato y se compararon con el spline de la estación de Jato.

Se obtuvo la siguiente gráfica:

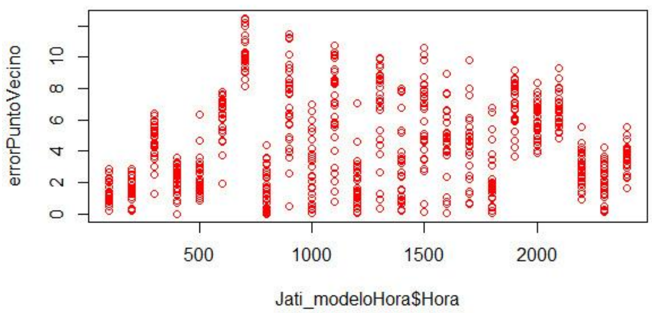

Ilustración 8 Error Vecino más Cercano

Se puede observar que la mayoría de los errores se encuentra entre 0 y 2, unas pocas entre 2 y 6, y un error único que supera el 16.7, este último puede ser que el spline en ese valor se equivocó al poder ser un valor máximo o mínimo, el cual va a inflar mucho el error medio.

- Error máximo: El error máximo se encuentra alrededor de la hora de 575 y tiene un valor de 16.7
- Error mínimo: El error mínimo es 0, ya que hubo una hora donde el valor del spline y el valor real son iguales.
- Error medio: Al sumar el valor máximo que es 16.2 y mínimo que es 0, y dividirlo entre 2, da como resultado
- Error absoluto: Al sumar todos los errores y dividirlos entre la cantidad de valores da como resultado 1.30, siendo un valor relativamente bajo.

De este resultado se puede decir que el spline de los valores de la estación Jato son valores cercanos a los valores reales, con un error promedio de 1.30, por tanto, aplicar el método del vecino más cercano es correcto para este caso.

AGRADECIMIENTOS

Para finalizar, agradecemos encarecidamente por el apoyo para la realización de este reto a nuestra profesora Eddy Herrera. La realización de este reto no habría sido posible sin su continuo apoyo, sobre todo, en los tiempos de incertidumbre tanto por temas sanitarios como por agitaciones políticas en las cuales nos vemos enrevesados. Su paciencia y dedicación fueron un factor determinante para mantener la motivación y la curiosidad dentro del equipo para la finalizar satisfactoriamente este reto.