Interpolación de temperaturas registradas por estaciones climáticas

Santiago Andrés Caroprese Hidalgo, Daniel Hernández García, Juan Carlos Suárez Jaimes

1Análisis Numérico, Ingeniería de sistemas, Pontificia Universidad Javeriana

**Resumen: En el presente documento se describe la aplicación de métodos de interpolación a partir de los datos registrados por estaciones climáticas para obtener temperaturas en instantes de tiempo ausentes en la muestra y para interpolar los datos de un punto geográfico a partir de sus estaciones cercanas. Se encontró que los métodos de Lagrange, Lagrange baricéntrico y Newton no funcionan correctamente en este tipo de problemas por la gran cantidad de datos involucrados, por lo que resulta necesario utilizar un enfoque “por partes” como spline cúbico o interpolación cúbica de Hermite por partes, entre los cuales el segundo presentó mejores resultados. En cuanto a la interpolación a partir de estaciones cercanas, se utilizó el método IDW, para el cual se encontró que el valor óptimo del parámetro potencia es menor que el usado normalmente, por lo que se concluyó que este puede depender en gran medida de las características del área geográfica analizada.**

1. Introducción

El conocimiento meteorológico ha sido de gran utilidad para el manejo óptimo de la agricultura, recursos naturales, anticipación y recuperación de desastres, urbanización, transporte, entre otras.

En el presente documento se usarán datos meteorológicos para realizar dos ejercicios de interpolación. En el primero se busca utilizar una muestra de datos de temperatura para interpolar, a través de varios métodos, la temperatura en instantes de tiempo ausentes en la muestra. En el segundo, se busca aproximar la temperatura en un punto geográfico a partir de los datos registrados en estaciones cercanas, a través del método de distancia inversa ponderada (IDW), con el objetivo de determinar el valor óptimo del parámetro potencia.

El artículo [1] expone un ejemplo de aplicación de la interpolación de datos entre estaciones climáticas, para estimar los mismos valores en diferentes zonas del planeta tierra. En el artículo se presenta un nuevo método de interpolación espacial modificado, de ponderación de distancia inversa para temperatura ambiental, precipitación, y humedad relativa integrando datos satelitales con datos de estaciones meteorológicas. El uso de interpolación espacial para predecir el estado meteorológico resulta de importancia, ya que el número de estaciones es limitado, teniendo en cuenta el área geográfica, pero esta información resulta de gran utilidad, tal y como se explicó previamente.

En el artículo se comparan diversos métodos, para validar los resultados obtenidos por el método planteado. Este presenta características de la interpolación de datos, en busca de un resultado óptimo, y el uso de pesos según las distancias entre las estaciones con el lugar objetivo. Este método planteado resulta ser una variación del método IDW, el cual es utilizado en el presente documento.

1. Desarrollo de Contenidos

La interpolación de Lagrange es usada para construir un polinomio con grado máximo n que pase por n+1 puntos (x0, y0), (x1, y1),…, (xn, yn).

Para Lagrange, es necesario construir para cada k=0,1, 2,…,n. una función Lk(x) llamada la base de Lagrange, con la propiedad que Lk(xi) es 1 cuando i=k y 0 cuando no, por lo cual está definida como [2]:

Luego un polinomio único de interpolación p(x) existe si n+1 puntos son dados, , p(x) está definido por:

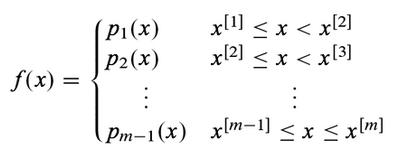
Otro método de interpolación es el de Newton. Este es parecido al de Lagrange, pero su base está definida por:

Existe otra versión del método de Lagrange, llamada Lagrange baricéntrico. Usando el concepto de peso baricéntrico, dado por la fórmula [2]:

Con lo que luego se obtendría la función de Lagrange baricéntrico que determina el polinomio [2]:

La interpolación por spline cúbico usa el enfoque de aproximación polinómica "por partes". Este enfoque se basa en dividir el intervalo de aproximación en una colección de subintervalos y construir un polinomio en cada subintervalo. Los splines cúbicos hacen uso de polinomios cúbicos para cada subintervalo. Se necesitan al menos 4 puntos [3].

Un spline cúbico es una función compuesta por polinomios cúbicos en m subintervalos:



Una función spline cúbico f(x), debe cumplir las siguientes condiciones [3]:

a) La función por partes f(x) debe interpolar todos los puntos en cada subintervalo [xi,xi+1].

b) f(x), f’(x), f’’(x) son continuas en a <= x <= b de manera que:

* 1. fi-1,i(xi) = fi,i+1(xi)
  2. f’ i-1,i(xi) = f’ i,i+1(xi)
  3. f’’ i-1,i(xi) = f’’ i,i+1(xi)

En la interpolación cubica de Hermite por partes, también se tiene el enfoque de dividir el intervalo [a,b] de interpolación en n subintervalos por lo que también debe cumplir las condiciones de spline cúbico, se diferencia de splines cúbicos en el sentido de que solo busca encajar las derivadas de primer orden en los puntos dados con los del intervalo antes y después. Su polinomio es representado por [4]:



Donde P0 es el punto inicial, P1 el punto final, m0 es la tangente inicial, m1 es la tangente final y h00, h10, h01, h11 son las funciones base de Hermite [5]:

Distancia inversa ponderada o IDW es un método determinístico para interpolación multivariada basada en el principio de que el valor de un punto estimado está más correlacionado con un punto conocido cercano que con uno lejano. Conforme la distancia entre puntos estimados y conocidos aumenta, la eficacia del punto medido disminuye. Este efecto es tenido en cuenta por el método, dado que los pesos utilizados son inversamente proporcionales a una potencia de la distancia [1]. La ecuación de IDW es la siguiente.

En esta ecuación, u0 es la ubicación en la que se está realizando la estimación y ui, i = 1 … N son las ubicaciones en las que se tienen mediciones conocidas. El valor estimado es un promedio estimado de N valores medidos en ubicaciones diferentes, i = 1… N [1]. Los pesos se calculan a través de siguiente ecuación.

En esta expresión, es la distancia euclidiana entre el punto estimado y el de la medición. Por lo general, el parámetro de potencia p toma el valor de 2, que reduce el peso de los puntos lejanos al momento de realizar la interpolación [1].

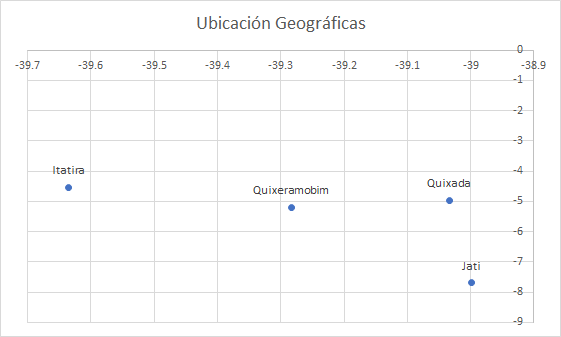
1. Métodos y Datos

En primer lugar, se llevó a cabo la interpolación de temperaturas a partir de una muestra. Para esto, se utilizaron 703 datos de temperatura interna (en ºC) tomados en diferentes instantes de tiempo en la estación Fortaleza UECE (Brasil), en el 2013. Los datos que se adaptaron para que iniciaran con t0=0, de tal manera que el tiempo fuera representado a través de una sola variable continua, ya que, inicialmente el tiempo se dividía en año, día y hora. La interpolación se realizó a partir de una muestra que representa el 80% de los datos con los que se cuenta, escogida de manera aleatoria. Una vez obtenido el polinomio interpolador, este se utilizó para interpolar los valores ausentes en la muestra, con el fin de comparar los valores interpolados con respecto a los reales, con el objetivo de calcular los errores correspondientes. Este proceso fue realizado utilizando diferentes métodos de interpolación: Lagrange, Lagrange baricéntrico, Newton, spline cúbico y Hermite cúbico por partes.

En segundo lugar, se realizó la interpolación de datos de una estación a partir de estaciones cercanas. Para esto, se utilizaron los datos de temperatura correspondientes a tres estaciones (Itatira, Jati y Quixada) para interpolar los datos de la estación de Quixeramobin, a partir del método IDW. De las diferentes estaciones que se tenían para elegir, se eligieron estas tres, debido a que son las únicas que poseen los 720 datos de la temperatura interna en el intervalo utilizado. Se consideró que tener una mayor cantidad de datos para realizar la interpolación permitiría obtener mejores resultados más acertados. Los resultados de la interpolación fueron comparados con los datos reales de Quixeramobin con el objetivo de calcular el error y, a partir de esto, obtener el valor óptimo de p.

|  |  |  |
| --- | --- | --- |
| Estaciones | Coordenada X | Coordenada Y |
| Quixada | -39.0334 | -4.967 |
| Jati | -38.9979 | -7.68238 |
| Itatira | -39.636 | -4.52894 |
| Quixeramobim | -39.2837 | -5.20077 |

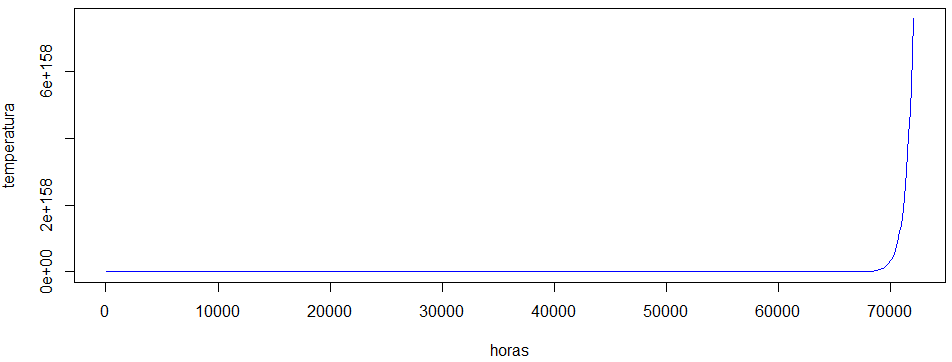
1. Tabla. Ubicación geográfica de estaciones utilizadas para interpolación espacial.



**Gráfica. 1.** Ubicación geográfica de estaciones utilizadas para interpolación espacial.

1. Resultados
   1. Interpolación de temperaturas a partir de una muestra

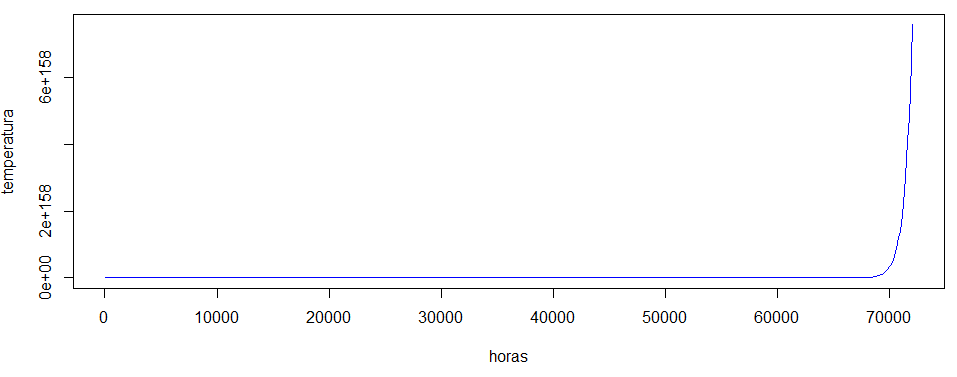
Con la interpolación de Lagrange, se obtuvieron los resultados mostrados en la gráfica 2. Se puede observar que los valores de la temperatura tienden a infinito, por lo que los resultados no son satisfactorios.



**Gráfica.2.** Temperatura interpolada con Lagrange vs. horas.

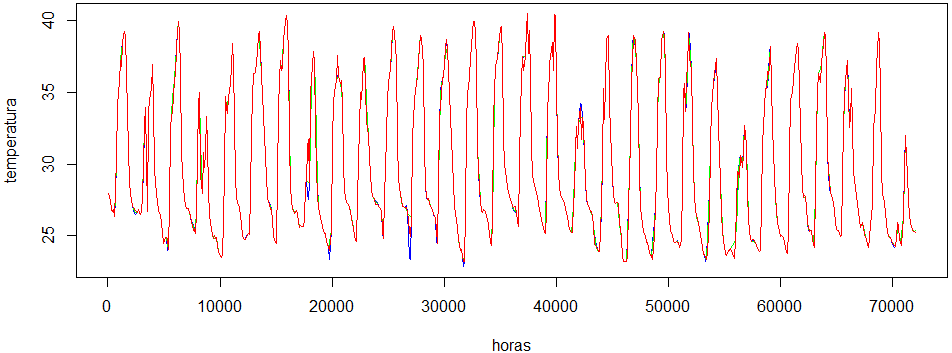
Al utilizar el método de Lagrange baricéntrico se observó que no era posible interpolar los puntos ausentes a partir de los puntos de la muestra. Más adelante se explicará el análisis realizado al respecto.

En cuanto al método de Newton, en la gráfica 3 se puede observar que presenta el mismo comportamiento observado para el método de Lagrange. Al igual que Lagrange, los valores de la temperatura tienden a infinito, por lo que los resultados no resultan satisfactorios.



**Gráfica.3.** Temperatura interpolada con Newton vs. horas.

Los métodos de spline cúbico y Hermite cúbico por partes, en cambio, sí permitieron interpolar los puntos ausentes a partir de los puntos de la muestra, los cuales fueron comparados. En la gráfica 4, se muestran graficadas la temperatura real en color rojo, la interpolación con spline cúbico en color azul y la interpolación de Hermite cúbica por partes en color verde. La temperatura real se graficó encima de las interpolaciones con el objetivo de que se evidenciaran de manera clara las partes en las que las interpolaciones no coinciden con el valor real. En la tabla 2 se muestran los diferentes tipos de error para estas dos interpolaciones.



**Gráfica.4.** Temperatura vs horas.

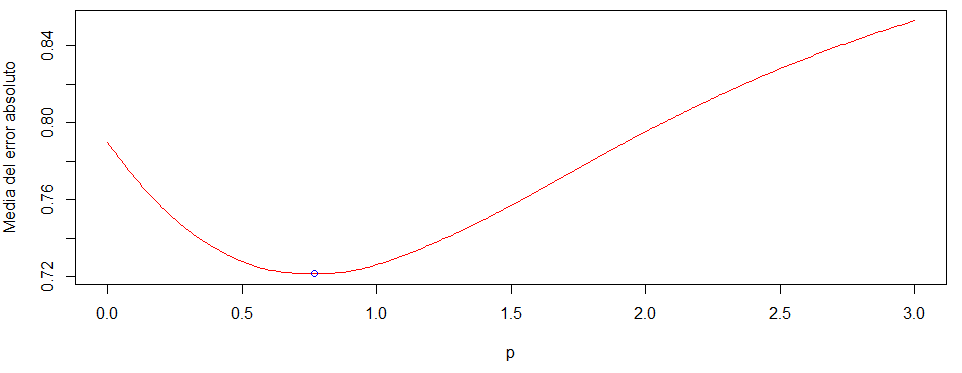
Temperatura real en color rojo, interpolación con spline cúbico en color azul e interpolación de Hermite cúbica por partes en color verde.

|  |  |  |
| --- | --- | --- |
| Método | Spline cúbico | Hermite cúbico por partes |
| Media de error absoluto | 0.6059463 | 0.6017391 |
| Error absoluto mínimo | 0.001832248 | 0.0006400839 |
| Error absoluto máximo | 4.187121 | 3.272722 |
| Índice de Jaccard | 0.6676428 | 0.6686217 |

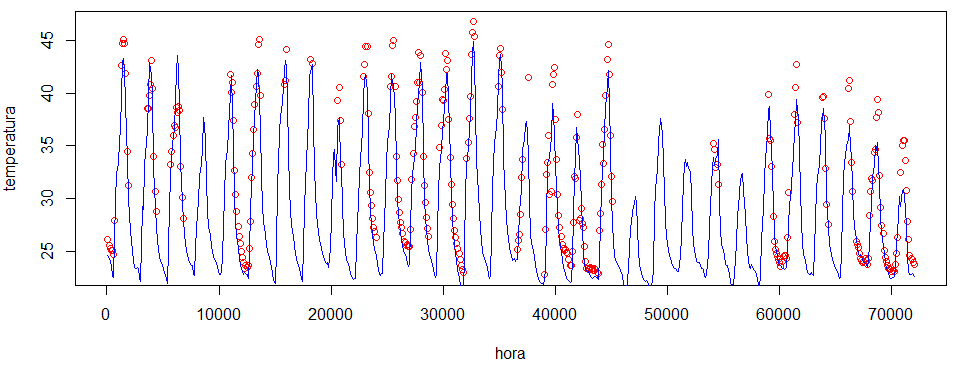
2. Tabla. Error de interpolación con spline cúbico e interpolación de Hermite cúbica por partes.

1. Interpolación de datos de una estación a partir de estaciones cercanas

Se aplicó el método IDW variando el parámetro potencia, utilizando valores entre 0 y 3, con paso de 0.01. Al graficar los errores obtenidos con cada uno de estos valores del parámetro potencia p, se obtuvo la gráfica 5. En esta, se marca también el punto en el cual el error es mínimo, que se produce en p=0.77. En la gráfica 6, se muestra la gráfica en la cual se puede apreciar la interpolación de la temperatura obtenida con este parámetro potencia óptimo. En la tabla 3, se muestran los valores del error para esta interpolación.



**Gráfica.5.** Media de error absoluto vs. parámetro potencia. Valor óptimo marcado en azul.



**Gráfica.6.** Temperatura vs. hora. En azul, se muestra la interpolación realizada con el método IDW, utilizando el valor de p óptimo, 0.77. En rojo, se muestran marcados los puntos reales.

|  |  |
| --- | --- |
| Parámetro potencia | 0.77 |
| Media de error absoluto | 0.7214029 |
| Error absoluto mínimo | 0 |
| Error absoluto máximo | 6.447164 |

3. Tabla. Error de interpolación con IDW utilizando el parámetro potencia óptimo.

1. Conclusión

En cuanto a la interpolación de temperaturas a partir de una muestra, se evidenció que varios de los métodos utilizados no funcionaron correctamente con los datos del problema en cuestión.

En primer lugar, se evidenció que en este caso no se puede utilizar Lagrange Baricéntrico porque los pesos baricéntricos son cero. El peso baricéntrico wj se calcula a partir de una expresión que tiene en el denominador una productoria de las diferencias entre xj y los demás x. Sin embargo, como en este caso se intenta realizar la interpolación a partir de alrededor de 500 puntos, donde muchos de ellos presentan diferencias muy grandes, el valor de esta productoria es demasiado grande, haciendo que el valor de la expresión tienda a 0. Esto provoca que luego se obtenga una expresión 0/0 al reemplazar los valores de los pesos baricéntricos. Por lo tanto, se puede concluir que este método no es aplicable en los casos en los que se quiera obtener un polinomio interpolador para una gran cantidad de puntos donde los valores de x presenten grandes diferencias.

En segundo lugar, se observó que tanto el método de Newton como Lagrange produjeron resultados en los cuales las temperaturas tienden a infinito. Al revisar los polinomios obtenidos a través de estos métodos, se observó que los coeficientes crecen gradualmente hasta tomar valores de valor absoluto demasiado grande, que se salen de los límites de la máquina, por lo que después de cierto punto, estos valores son guardados como cero. Es importante tener en cuenta que este conjunto de datos requería obtener un polinomio de grado 500, aproximadamente, por lo que se puede concluir que estos métodos no son apropiados para realizar interpolaciones a partir de muestras tan grandes.

Teniendo en cuenta que se determinó que no es conveniente calcular un solo polinomio de interpolación para muestras de datos tan grandes, resulta más conveniente partir los datos de la muestra en intervalos y generar un polinomio interpolador para cada uno de estos intervalos. De hecho, esta es la metodología utilizada por los dos métodos que permitieron interpolar los datos sin tener los problemas descritos anteriormente: spline cúbico y Hermite cúbico por partes. Al comparar ambos métodos en la gráfica 4, se puede observar las interpolaciones de los puntos ausentes en la muestra son bastante cercanas la mayoría de las veces. Sin embargo, resulta evidente que la interpolación con spline cúbico tiende a alejarse un poco en los máximos o mínimos locales, mientras que Hermite cúbico por partes no suele tener este problema. Por otro lado, se puede observar que, en las partes en las que no hay picos, la interpolación con spline cúbico suele ser más acertada, siendo tapada completamente por los valores originales, mientras que Hermite cúbico por partes, si bien se acerca bastante, en ocasiones se aleja un poco más, por lo que llega a ser visible en la gráfica.

Al observar los datos del error, se evidencia que Hermite cúbico por partes presenta menores valores de media del error absoluto, error absoluto mínimo y error absoluto máximo que spline cúbico. Hermite cúbico por partes presenta, a su vez, un índice de Jaccard ligeramente mayor. A partir de esto, se puede concluir que, para este caso, la interpolación de Hermite cúbico por partes fue más acertada, pero, a partir de la gráfica, se considera que esto se debe a que la gran cantidad de picos en los datos provocó que aumentara el error de spline cúbico. Por lo tanto, se puede concluir que la interpolación de Hermite cúbico por partes es la mejor opción para problemas como este en el que hay gran cantidad de picos en los datos, pero en problemas con menos picos es probable que spline cúbico presente mejores resultados.

En cuanto a la interpolación de datos de una estación a partir de estaciones cercanas, en la gráfica 5 se puede observar que el parámetro potencia tiene un gran impacto sobre el error de la interpolación del método IDW. Resulta especialmente llamativo que, para los datos utilizados, el valor óptimo del parámetro potencia sea 0.77, que se encuentra bastante por debajo del valor usualmente utilizado, que es 2. El reducir el valor del parámetro potencia, provoca que se decrementen en menor medida los pesos de las estaciones lejanas. Es decir, con este valor de p, las estaciones cercanas y lejanas tienen pesos más similares. Por lo tanto, se puede concluir que, en este caso, el área geográfica cuyos datos fueron analizados no presenta una variabilidad tan grande en la temperatura, por lo que incluso estaciones lejanas tendrán temperaturas algo similares. Sin embargo, se considera que el valor del parámetro potencia puede depender en gran medida de las características del área geográfica que se esté analizando, por lo que un área geográfica con alta variabilidad en sus temperaturas podría requerir utilizar valores de p mucho mayores que el que resultó ser el óptimo para este caso.

En la gráfica 6 resulta evidente que la interpolación falla, sobre todo, en los puntos máximos. Se puede observar que, en estos puntos, el valor interpolado resulta menor que el real, llegando a un error absoluto máximo de 6.447164. Sin embargo, resulta llamativo que, en el resto de los puntos, la interpolación resulta bastante acertada. Como resultado de esto, la media del error absoluto se mantiene en 0.7214029. A partir de esto, se puede concluir que, si bien el comportamiento general de la temperatura puede ser interpolado de manera bastante acertada a través de este método, los valores máximos pueden llegar a representar cierta dificultad. Sin embargo, se considera que esto puede depender en gran medida de las características del área geográfica analizada y de las estaciones utilizadas para la interpolación, dado que es probable que con una muestra más grande pueda lograrse utilizar la información de estaciones con máximos más similares para reducir este error. En este caso particular, se considera que, para las horas en las que no se tienen datos de la estación de Quixeramobin, la interpolación podría ser usada para obtener una estimación bastante acertada de la temperatura, siempre y cuando se trate de horas en las cuales no se producen los máximos.

1. Referencias

|  |  |
| --- | --- |
| [1] | E. Ozelkan, S. Bagis, E. Ozelkan, B. Ustundag, M. Yucel and C. Ormeci, “Spatial interpolation of climatic variables using land surface temperature and modified inverse distance weighting.,” *Internation Journal of Remote Sensing,* vol. 36, no. 4, pp. 1000-1025, 2015. |
| [2] | J. P. Berrut and L. n. Trefethen, “Barycentric Lagrange Interpolation,” *Society for Industrial and Applied Mathematics,* vol. 46, no. 3, pp. 501-507, 2004. |
| [3] | L. A. Leal, “Towards data science,” 2 Diciembre 2018. [Online]. Available: https://towardsdatascience.com/numerical-interpolation-natural-cubic-spline-52c1157b98ac. [Accessed 12 Noviembre 2020]. |
| [4] | A. Pierce., “University of British Columbia.,” 15 Septiembre 2012. [Online]. Available: http://www.math.ubc.ca/~peirce/M406\_2012\_Lecture\_4n5.pdf. [Accessed 12 Noviembre 2020]. |
| [5] | N. Pipenbrick, “cubic,” 30 Marzo 1998. [Online]. Available: https://www.cubic.org/docs/hermite.htm. [Accessed 12 Noviembre 2020]. |