



Institución : Universidad Nacional Autónoma de México
Sede: Instituto de Investigaciones en Matemáticas Aplicadas
y en Sistemas
Curso : Matemáticas Discretas
Fecha : 3 de diciembre de 2021
Autor : Ayala Salcedo Marlene Viviane; Fuentes Cruz Luis
Ariel
Mail : marleneayala@ciencias.unam.mx ; arielfuen-
tes@ciencias.unam.mx

Reporte ejecutivo

1. Planteamiento del problema a resolver

El Sistema de Transporte Colectivo Metro es posiblemente el principal medio de transporte público utilizado en la Ciudad de México. Por lo tanto, es uno de los lugares más potenciales donde se puede sufrir un contagio del virus *SARS – COV – 2*, (COVID-19), debido a que es una enfermedad que se propaga por vías respiratorias, (es decir, nariz y boca). Más aún, dicho virus se propaga con mayor facilidad en espacios interiores o en aglomeraciones de personas.

Al ser esta una enfermedad relativamente nueva, los países no cuentan con suficiente información para el tratamiento de la misma, por lo que las principales recomendaciones son permanecer en casa, mantener una distancia de aproximadamente 16 pies entre cada par de personas, utilizar cubre bocas, así como evitar aglomeraciones.

Pese a que el índice de muertes en México se ha visto reducido y hasta controlado, no lo es así el número de casos nuevos registrados de la enfermedad. Es por ello que debería seguir siendo una prioridad, si no el mantenerse en casa, sí el evitar aglomeraciones y por supuesto, seguir las medidas sanitarias como el lavado de manos y uso del cubrebocas.

Con el fin de ofrecer rutas con el menor índice de afluencia, y por tanto de aglomeración de personas, el objetivo del presente proyecto es hallar la ruta, dadas una estación inicial, una estación final y un día de la semana, que sea menos transitada. El fin de esto es intentar minimizar las aglomeraciones sufridas dentro de la red del Sistema de Transporte Colectivo Metro, evitando así posibles contagios.

Otra de las ventajas de las rutas con menor afluencia es que, como se sabe, las estaciones más concurridas suelen significar un traslado más lento de un lugar a otro, por lo que encontrar una ruta con menos afluencia de personas no sólo disminuye el contagio, sino también puede ser clave para disminuir el tiempo de traslado.

Luego entonces queda definido el objetivo del presente proyecto; implementar un programa que ayude a los usuarios a obtener la ruta con menos tránsito de personas, de manera que no deban comprometer su salud ni su tiempo para trasladarse de un lugar a otro a consecuencia de la cantidad de gente que utiliza el servicio del Sistema de Transporte Colectivo Metro diariamente.

2. Justificación de elección de datos base y complementarios

Para lograr el objetivo, se emplearon dos bases de datos. La primera y principal es la base de datos *Afluencia diaria del Metro de CDMX* obtenida de la página oficial de datos del gobierno de la Ciudad de México. Dicha base contiene la fecha del registro, una columna con el día, una con el mes y una con el año. Además, contiene la estación, línea del metro al que corresponde la estación y finalmente la afluencia del día del registro para la estación correspondiente.

La segunda base de datos fue la base de datos *Coordenadas*, extraída del portal oficial de datos de la Ciudad de México. Esta contiene la ubicación geográfica de cada una de las estaciones del metro, así como el nombre de la estación.

De la primera base de datos, la información proporcionada es clave para poder obtener la afluencia promedio por estación, que depende del día de la semana, con el fin de obtener la ponderación de nuestras aristas (conexiones) que se utilizarán para la creación de la red del Metro. En la implementación, se utilizaron las columnas fecha, línea, estación y afluencia, además de algunas generadas posteriormente.

La segunda base de datos se utilizó para poder construir de manera gráfica la red del Metro y para representar en ella el camino con menor afluencia de una estación inicial, u , a una estación final v , dadas por el usuario. Para lo anterior, se utilizaron las columnas de latitud, longitud y el nombre de la estación.

3. Metodología descrita en términos no técnicos

3.1. Limpieza de datos

En términos generales, la primera base de datos requería un extenso tratamiento de los datos; se eliminaron las columnas que resultaban redundantes para el análisis, se corrigieron los nombres de las estaciones para tener un estándar de ellos, se realizaron cambios en el tipo de variable de algunos de los datos, pues teníamos varias inconsistencias entre los tipos de datos. Finalmente se agregó la columna que identificaba el día de la semana de cada registro.

3.2. Metodología del proyecto

Tras haber limpiado, transformado y cargado los datos de ambas bases, se prepararon los datos para la aplicación del algoritmo utilizado. En este proceso se calcularon las afluencias promedio de cada estación por día de la semana, de manera que para cada línea del metro se tenía la cantidad promedio de personas que utilizaban dicha estación dado un día de la semana.

Luego fue calculada la afluencia promedio por día de la semana bajo las afluencias diarias, mismas que se registraron en una matriz de 163×163 , 163 son el número total de estaciones diferentes que hay en la red del metro, sin contar transbordos. De esta manera, cada entrada de la matriz tenía un 0 si la estación i no estaba directamente conectada con la estación j , y la afluencia de una estación destino si dos estaciones estaban conectadas.

Se tomaron direcciones para las conexiones por lo que si i estaba conectada con j , también se cumplía que j estaba conectada con i . Además, la entrada i, j de la matriz anteriormente descrita, tenía la afluencia del nodo j y la entrada j, i tenía la afluencia del nodo i .

Después de obtener la matriz, se creó la gráfica correspondiente a partir de ella.

Una vez obtenida dicha gráfica se utilizó el algoritmo de Dijkstra modificado.

Ahora, antes de hablar del algoritmo de Dijkstra modificado, daremos la idea general detrás del algoritmo de Dijkstra, mismo que consiste en ir explorando todos los caminos más cortos que parten del vértice origen y que llevan a todos los demás vértices; cuando se obtiene el camino más corto desde el vértice origen hasta el resto de los vértices que componen el grafo, el algoritmo se detiene. Por lo tanto el uso de este algoritmo nos sirve para encontrar el camino más corto de un nodo a otro dado, (algoritmo de Dijkstra modificado), en donde entenderemos por camino más corto al camino con menor afluencia de gente de una estación a otra.

Finalmente se trazó sobre la red del metro el camino de la estación inicial a la estación final.

4. Discusión no técnica de uso de recursos computacionales

A diferencia del algoritmo original de Dijkstra, el algoritmo modificado empleado únicamente se centra en el nodo inicial, correspondiente a una estación inicial y el nodo final, que alude a una estación final. Por esto, el tiempo de ejecución es más rápido que si utilizáramos el algoritmo original y devolviéramos sólo el camino de la estación inicial a la estación final. Además al no guardar todos los caminos de la estación inicial al resto de las estaciones, se optimiza el uso de espacio en disco, así como los tiempos de ejecución.

A lo largo de la implementación se pueden ver funciones auxiliares, mismas que fueron implementadas con el fin de ser lo más óptimas posibles.

Respecto a los recursos computacionales, debemos resaltar que la base de datos principal cuenta con cerca de 45 mil registros compuestos por 8 características, teniendo un peso de 2,2 Mb, que es relativamente bajo, por lo cual el uso de memoria no se ve comprometido mayormente. Sin embargo, la transformación de los datos sí fue, temporalmente, costosa, pues debido a la cantidad de datos, un proceso en particular (correspondiente a la determinación de afluencias), no se ejecutó tan rápido en comparación con el resto de llamadas a funciones y procesos.

Sobre la segunda base de datos, el tamaño de esta es mucho menor, por lo cual tampoco se compromete a los recursos computacionales.

El *output* de la función final es un *GIF*. En las llamadas a la función principal empleadas como prueba el tamaño promedio de los GIFs fue de 6 Mb, por lo cual los resultados tampoco podrían llegar a comprometer los recursos de memoria del sistema. Por otro lado, al ser una animación el resultado final, los tiempos de ejecución son claramente mayores a los que registraría un programa análogo que sólo regresara la lista de estaciones a seguir para llegar de una estación a otra.

5. Conclusiones

Tras la implementación detallada anteriormente y tras probar la implementación, se concluye que;

- El objetivo principal, (encontrar la ruta menos afluyente dadas dos estaciones pertenecientes a la red del Sistema de Transporte Colectivo Metro de la CDMX y un día de la semana), se logró con satisfacción, proporcionando así alternativas al traslado de personas que signifiquen un riesgo menor de contagio y, potencialmente, bajo un tiempo de traslado menor.
- Tras analizar las afluencias promedio, y como conclusión secundaria, se observó que las estaciones con más afluencia son aquellas que fungen como transbordos.
- Como propuestas y trabajos futuros, se propone la implementación de un algoritmo de memoización para determinar los caminos menos afluentes entre cualesquiera dos estaciones en cualquier día. Esto, podría ser implementado corriendo el algoritmo presentado en todos los pares de nodos o con el algoritmo de Floyd-Warshall.
- Se propone dar seguimiento a la base de datos principal con el fin de que los resultados fueran mucho más precisos. Más aún, se propone la implementación de una base de datos con la afluencia diaria tomando en cuenta el horario. Esto, significaría resultados mucho más realistas.