GoMotion: intensidad de movilidad y detección de picos

Javier Badesa
javierbadesa57@gmail.com

Alexander C. Hall
alexanderhallparrilla@gmail.com

Oscar Senesi
oscarsenesi@protonmail.com

Dan M. Vancea
danvancea1235813@gmail.com

Índice

1. Introducción

Para estudiar y mejorar el transporte en la ciudad de Barcelona, es necesario conocer la disposición e infraestructura que hay en la localidad. Por ello, el objetivo de este documento es crear una abstracción matemática de la ciudad que permita modelar las propiedades reales a partir de un conjunto de datos proporcionados.

2. Definiciones, restricciones y suposiciones

2.1. Conjunto de datos

Antes de proceder con el análisis, primeramente hemos de destacar qué atributos deben tener nuestros datos. A continuación se supone que se tienen datos estructurados en filas, y cada una de ellas tiene como mínimo:

- Un punto de salida
- Un punto de llegada
- La cantidad de personas desplazadas entre ambos en un periodo determinado de tiempo (T)

En nuestro caso, tomaremos datos con frecuencia diaria y barrios como puntos.

2.2. Ciudad en barrios

Nuestro modelo considera la ciudad como un conjunto de barrios conectados entre sí, por lo que es natural formalizar esto usando un grafo.

Por tanto, a partir de ahora consideramos un grafo $G = (V, E)$, donde V es el conjunto de vértices (que representan barrios) y E es el conjunto de aristas del grafo: X e Y están unidos por una arista si los barrios son adyacentes (comparten una frontera física), y la arista tiene un peso $e_{XY} = d(X, Y)$. Para la distancia entre dos barrios adyacentes utilizamos la distancia entre sus puntos medios. Para facilitar la lectura, usamos $V_X \subset V$ para denotar el conjunto de vértices con los que $X \in V$ comparte una arista.

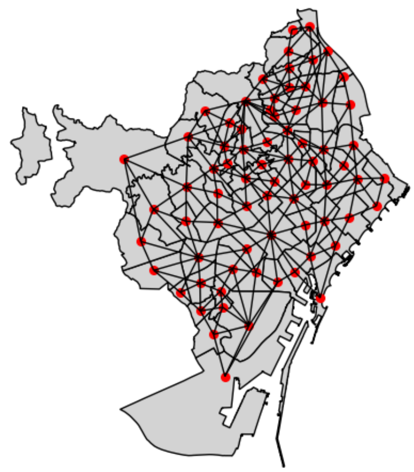

Figura 1: Visualización del grafo de Barcelona

Además, utilizaremos la notación N^{AB} para referirnos a el número de personas que se han desplazado de A a B registrado en los datos.

2.3. Intensidad de movilidad

Definimos la intensidad de movilidad como el número de desplazamientos entre 2 puntos en un periodo de tiempo determinado. Particularmente, la intensidad entre barrios adyacentes $\gamma(X, Y)$ es el número de personas que se han movido de uno al otro en un día (ajustamos la precisión de la capacidad al menor periodo que permiten los datos).

2.4. Restricciones sobre el desplazamiento

Para conseguir un modelo completo y con valor analítico, debemos tener ciertas restricciones sobre el desplazamiento de las personas. Supondremos que para llegar desde el origen hasta el destino se toma la trayectoria cuya longitud es mínima.

3. Modelo

3.1. Trayectorias

Una trayectoria es la sucesión de puntos de los que consiste un desplazamiento. Denotaremos una trayectoria que lleva de A a B como $t^{AB} = (t_0^{AB}, t_1^{AB}, \ldots, t_n^{AB})$ donde $t_i^{AB} \in V$. Claramente, t^{AB} tiene $n + 1$ elementos, $t_0^{AB} = A$ y $t_n^{AB} = B$.

La restricción sobre las trayectorias es que $(t_i, t_{i+1}) \in E$, pues la sucesión debe tener una interpretación real (no podemos movernos entre dos barrios que no sean contiguos).

Usaremos S como el conjunto de todas las trayectorias reales. Una trayectoria es real si un individuo la recorrería para llegar de A a B. Es decir, $s^{AB} \in S$ si su longitud

$$l(s^{AB}) = \sum_{i=1}^{n} d(s_{i-1}^{AB}, s_i^{AB}) = \sum_{i=1}^{n} e_{s_{i-1}^{AB} s_i^{AB}}$$

es mínimo sobre el conjunto de trayectorias $\{t^{AB}\}$. Dada la relación con los pesos de las aristas, podemos hallar las trayectorias reales fácilmente con el algoritmo de Dijkstra.

Notamos que $l(s^{AB}) \neq d(A, B)$ en la mayoría de los casos, pues esto simularía las distancias como las de un simple plano euclídeo, eliminando el criterio de selección y abstrayendo excesivamente la disposición de la ciudad.

3.2. Intensidad de una trayectoria

Análogamente a la definición entre barrios, la intensidad de una trayectoria t^{AB} (denotada por $\gamma(t^{AB})$) es el número de personas que recorren la trayectoria en un periodo T.

3.3. Intensidades de movilidad

En este apartado deduciremos las intensidades de movilidad entre barrios adyacentes. Para ello, estableceremos las ecuaciones que añaden los datos y determinan nuestro modelo.

Modelando la población como una variable continua,

$$N^{AB} = \int_0^T \gamma(s^{AB}) dt = \Gamma(A, B)T$$

donde Γ es la intensidad media en el periodo T.

Esta igualdad nos permite obtener las intensidades medias de todas las trayectorias, y con esta información podemos inferir las intensidades entre barrios adyacentes basándonos en una propiedad clave del modelo: todo desplazamiento entre 2 barrios adyacentes X e Y pertenece a una trayectoria que se está recorriendo, por lo que es la suma de todos los desplazamientos de las trayectorias que pasan por X e Y.

Definimos una función $\mathbf{I_s}(X, Y)$ que toma el valor 1 si existe un i tal que $s_i = X$ y $s_{i+1} = Y$ y 0 en el resto de casos. En otras palabras, esta función indica si el tramo $X \to Y$ es parte de la trayectoria s^{AB}. Luego la intensidad media entre 2 barrios X, Y conectados es

$$\Gamma(X, Y) = \sum_{s \in S} \Gamma(s)\mathbf{I_s}(X, Y) = \frac{1}{T} \sum_{(A,B) \in V \times V} N^{AB} \cdot \mathbf{I_{s^{AB}}}(X, Y)$$

4. Refinamiento del modelo

Para establecer el modelo, hemos supuesto que toda persona tomaría el camino más corto. Aunque esta es una aproximación bastante realista, no siempre es el caso que se da: puede haber gente que tome una pequeña variación, o más de una trayectoria con longitud cercana a la mínima. Además, es posible que hayan pequeñas variaciones locales de la intensidad de movimiento. Por ello, la existencia de más de una muestra de los datos nos da la oportunidad de crear un modelo probabilístico.

4.1. Múltiples trayectorias

Definimos $\Omega^{AB} = \{\omega_j^{AB}\}$ como el conjunto de las β trayectorias más cortas que van de A a B. β es un parámetro que determina el refinamiento del modelo, pues cuanto mayor es, más trayectorias posibles consideraremos. Computacionalmente se comprueba que $\beta \geq 20$ es suficiente para incluir todas las trayectorias relevantes. Para poder extraer conclusiones computacionalmente asequibles, afirmamos que una persona únicamente se desplazará de A a B siguiendo una trayectoria en Ω.

Con esta base, interpretaremos que el camino escogido por una persona desde A hasta B (denotado por ψ^{AB}) es ω_j^{AB} con probabilidad $p_{\omega_j^{AB}}$, en cuyo caso la intensidad entre barrios adyacentes X e Y

obtenida por el modelo se convierte en una variable aleatoria. Calculamos su valor esperado $\mathcal{I}(X,Y)$, que dado un gran número de viajes converge al valor real

$$\begin{aligned}
\Gamma(X,Y) \approx \mathcal{I}(X,Y) &= \mathbb{E}\left[\sum_{(A,B)\in V\times V} \Gamma(\psi^{AB}) \cdot \mathbf{I}_{\psi^{\mathbf{AB}}}(X,Y)\right] \\
&= \sum_{(A,B)\in V\times V} \left(\sum_{\omega\in\Omega^{AB}} p_\omega \Gamma(\omega) \mathbf{I}_\omega(X,Y)\right) \\
&= \frac{1}{T} \sum_{(A,B)\in V\times V} \left(\sum_{\omega\in\Omega^{AB}} p_\omega N^{AB} \mathbf{I}_\omega(X,Y)\right)
\end{aligned}$$

El desafío que presenta esta interpretación es encontrar la distribución de probabilidad de ψ^{AB} (que estimaremos a partir de su significado).

Para aproximar las distribuciones de probabilidad usamos las siguientes condiciones:

- La probabilidad de elegir una trayectoria depende de su longitud: cuanto más larga, menor ha de ser la probabilidad
- Dos trayectorias de igual longitud deben tener la misma probabilidad de ser escogidas
- Cualquier trayectoria cuya longitud se aleje suficientemente de la longitud mínima debe tener probabilidad despreciable, pues no sería racional elegir una trayectoria considerablemente más larga

Una función que permite modelar la proporción de forma satisfactoria es

$$p_{\omega_j^{AB}} \propto e^{-\alpha \frac{l(\omega_j^{AB})}{l(s^{AB})}}$$

para un parámetro $\alpha > 0$. Al ser una función decreciente de la longitud cumple los primeros dos puntos, y podemos escoger una α que represente la toma de decisiones de una persona promedio. Encontramos que $\alpha \approx 5$ elimina prácticamente las trayectorias más ineficientes manteniendo variedad en las trayectorias cercanas a la óptima (como referencia, $\alpha = 4{,}605$ otorga una probabilidad 100 veces menor de escoger un camino el doble de largo que el mínimo).

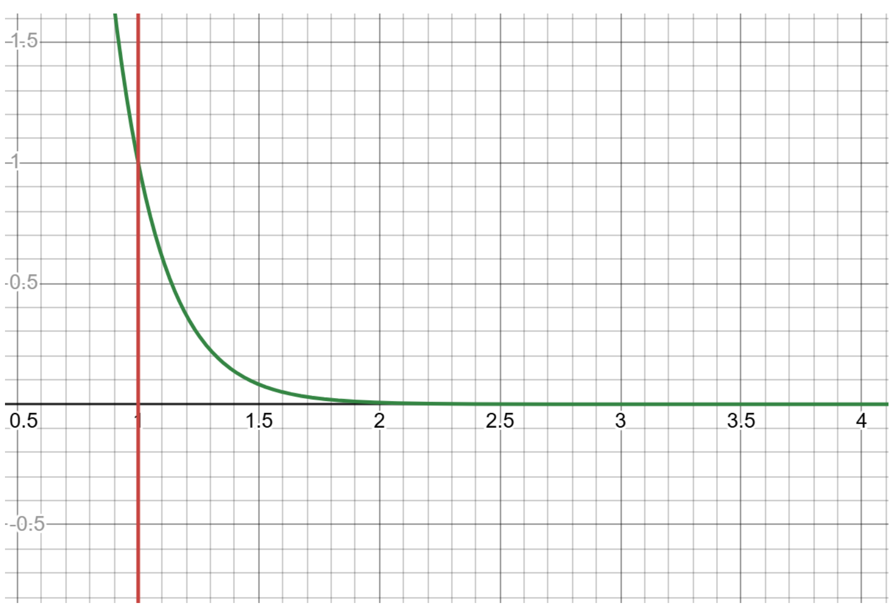

Figura 2: Probabilidad relativa al mínimo según $\frac{l(\omega_j^{AB})}{l(s^{AB})}$ para $\alpha = 5$

4.2. Extensión a varias muestras

Puesto que nuestro conjunto de datos contiene más de una muestra, somos capaces de crear herramientas y criterios que hagan uso de ellas.

Sea M el número de muestras que tenemos. El número de desplazamientos del barrio A al barrio B en la m-ésima muestra se escribe como N_m^{AB}, y de forma análoga todo subíndice m hace referencia a que el valor ha sido obtenido a partir de los datos de la m-ésima muestra.

4.3. Detección de picos

Una vez tenemos una metodología para estimar la intensidad entre barrios adyacentes, podemos utilizar métodos estadísticos convencionales para hallar valores que destacan y tomar decisiones sobre las medidas que se han de implementar. Una herramienta de análisis útil es el flujo de un barrio, que definimos de la siguiente manera:

$$\Phi_m(X) = \sum_{Y \in V_X} \Gamma_m(X,Y) + \Gamma_m(Y,X)$$

Utilizando el flujo podemos averiguar qué barrios actúan como focos de movilidad y qué tan intenso es el intercambio de personas.

- Pico de movilidad: $\Phi_m(X) \geq \sigma_{\Phi(X)} + \overline{\Phi(X)}$
- Pico masivo: $\Phi_m(X) \geq \frac{3}{2}\sigma_{\Phi(X)} + \overline{\Phi(X)}$

Otra magnitud que se obtiene a partir del modelo es el flujo total, una medida general de la movilidad en la ciudad:

$$\varphi_m = \sum_{(A,B) \in V \times V} \Gamma_m(A,B) = \frac{1}{2} \sum_{X \in V} \Phi_m(X)$$

Usando φ_m podemos detectar picos inusuales en un barrio teniendo en cuenta cuántas desviaciones estándar más se aleja un barrio de su media comparado con la ciudad entera. Es decir, existe un pico inusual si

$$\frac{\Phi_m(X) - \overline{\Phi(X)}}{\sigma_{\Phi(X)}} \geq \frac{\varphi_m - \overline{\varphi}}{\sigma_\varphi} + 1$$