

Integración y análisis de datos de movilidad escolar

Debido a que este Trabajo de Fin de Máster fue realizado con datos de estudiantes menores de edad y obedeciendo con lo estipulado en la Ley Orgánica de Protección de Datos, se considera conveniente no publicar. A razón de esto se opta por elaborar este resumen que contiene una explicación menos detallada del trabajo realizado.

En este trabajo se analizan los datos obtenidos a partir de una encuesta sobre movilidad escolar aplicada en distintos centros educativos del territorio gallego con el objetivo de construir indicadores de calidad de los caminos recorridos por los alumnos. Mediante encuestas geolocalizadas (proyecto GEOMOVE; <http://cartolab.udc.es/geomove>), se lograron recopilar datos periódicamente sobre el desplazamiento de alumnos a los distintos centros escolares. Este trabajo cumple con el objetivo de la representación de la dificultad en mapas, para facilitar la identificación de rutas óptimas hacia cada centro escolar, junto con la descripción de los principales obstáculos asociados a cada ruta, finalizando con la implementación de una aplicación web que permite obtener los datos resumidos de forma automática. En este trabajo se emplean principalmente técnicas para el análisis de procesos geoestadísticos.

Las cuestiones abordadas en esta encuesta se pueden agrupar en cuatro tipos: características del encuestado, hábitos de movilidad, percepción del recorrido y datos cartográficos. De camino al cumplimiento de los objetivos, los resultados arrojaron las conclusiones presentadas a continuación.

Al realizar el cálculo de la densidad de las localizaciones de los hogares de los alumnos, con la función **kde()**, observamos que generalmente se concentra una mayor densidad de hogares en los entornos del centro de estudio. Las medidas de resumen sobre las distancias hogar-colegio fueron calculadas a partir de la ruta más corta obtenidas con la función **dodgr_dists()** del paquete **dodgr** utilizando las geometrías geográficas recolectadas en la encuesta. Con estas herramientas podemos resumir que el 75% de los encuestados vive a menos de 3km de su respectivo colegio, pudiendo considerarse una distancia prudente para recorrer a pie.

Los resultados explorados referentes a los primeros tópicos tratados en la encuesta, muestran que pese a que los alumnos prefieren alternativas que favorecen al estado del tráfico y del ecosistema, 55.5% suelen ir en coche, siendo este un medio de transporte que sólo es preferido por el 17.7% de los encuestados.

Dentro de las respuestas sobre la percepción del recorrido de aquellos alumnos que suelen viajar en coche predominan las malas prácticas de los demás conductores al momento de aparcar. Alrededor del 45% de los encuestados que van caminando al colegio refieren que hay semáforos en los cruces de su recorrido. Al hablar de los hábitos, sólo el 43.2% cruza pasos peatonales mientras que el 69.6% camina por aceras. Las opciones de respuestas a estas preguntas puede que hayan condicionado la opinión de los alumnos en las próximas preguntas de la encuesta.

Se evalúan los resultados obtenidos en las preguntas abiertas sobre los elementos que mejorarían en el recorrido y la descripción de otros problemas en la ruta, implementando herramientas para la exploración de texto, observamos que las respuestas son bastante parecidas a las opciones que enlistan las preguntas iniciales de la encuesta.

Luego de realizar un análisis de correspondencias múltiples con el cual enmarcamos, desde un punto de vista descriptivo, la relación entre distintos grupos de variables, entre ellas resultando visible la relación

entre los alumnos de la ESO y los que perciben el trayecto al colegio como agradable, también podemos decir que se relacionan los alumnos que marcan el trayecto como desagradable y los que suelen ir en coche al centro de estudio. Consideramos que adicional a este análisis descriptivo se podría evaluar, futuros trabajos, la creación de clusters de alumnos para explorar los datos según la percepción del entorno, hábitos en el trayecto y modo de transporte usual.

Una vez marcado cada problema en el trayecto hacia el colegio se pide al alumno identificar, dentro de un listado, el tipo de problema en cada punto señalado. De no encontrarse el problema en el listado, se le solicita al estudiante que redacte una breve descripción del percance en la ruta. Al realizar un análisis de conglomerado jerárquico sobre estas descripciones, obtuvimos como resultado que sólo dos de ocho conglomerados se refieren a problemas diferentes a los enlistados dentro de las opciones de la variable *tipo de problema*.

Para estimar la intensidad tipo kernel de los problemas marcados por los alumnos utilizamos el estimador de difusión que ejecuta un algoritmo de diferencia finita, un novedoso método desarrollado por Mcswiggan et al. (2017)¹ para obtener intensidades de observaciones dentro de una red, implementado en el paquete **spatstat** del software R. Si bien es cierto que según las demostraciones de Mcswiggan et al. (2017) este estimador cuenta con mejores propiedades que las técnicas usualmente utilizadas, en algunas calles la intensidad calculada es alta sin tener un gran cúmulo de puntos cercanos y estando alejadas de los puntos de mayor intensidad. Al comunicarnos con los desarrolladores del paquete nos informaron que aún están tratando de descubrir cuál podría ser el problema, con la suposición de que el percance surge cuando la red tiene algunos segmentos que son muy cortos (en comparación con otros segmentos).

Finalmente se desarrolla una aplicación Shiny en R con el objetivo de obtener información de manera automática sobre los métodos aplicados en el trabajo para analizar los datos de la encuesta, en especial sobre las rutas seguidas los problemas marcados por los alumnos en cada colegio.

¹ McSwiggan, G., Baddeley, A. y Nair, G. (2017). Kernel density estimation on a linear network. *Scandinavian Journal of Statistics*, 44(2), 324-345.