

USO DE DATOS ABIERTOS PARA PRONOSTICAR Y MODELAR LOS TIEMPOS DE DESPLAZAMIENTO EN LA CIUDAD DE BOGOTÁ

USING OPEN ACCESS DATA TO FORECAST AND MODELING TRAVEL TIME IN URBAN BOGOTÁ

Fabián Castaño¹, Andrés Felipe Gutiérrez Bonilla², Ciro Amaya³, Nubia Velasco⁴

^{1,2,4}Universidad de los Andes, Facultad de Administración, Bogotá, Colombia

³Universidad de los Andes, Departamento de Ingeniería Industrial, Bogotá, Colombia

Resumen. La congestión vehicular constituye uno de los factores más importantes que afectan negativamente la calidad de vida en las grandes ciudades, tiempos de recorrido y, como consecuencia de esto, los procesos de distribución urbana de mercancías. No considerarlos puede implicar la creación de rutas ineficientes, desperdiciando recursos, incrementando la polución y disminuyendo los factores de carga de los vehículos. Este trabajo aborda el pronóstico y modelaje de tiempos de recorrido considerando la hora del día y la dirección del movimiento para transportarse entre pares de ubicaciones en la ciudad de Bogotá. Se plantea la utilización de regresiones mediante el método de los k-vecinos más cercanos para explotar la información consolidada, libremente disponible, ofrecida a través del sitio web Uber Movement ®. Posteriormente se propone la aproximación de los tiempos de desplazamiento mediante una función lineal por partes construida a partir de las predicciones. La información es utilizada para resolver un problema de ruteo de vehículos con tiempos de desplazamiento dependientes de la hora. Los resultados preliminares son comparados contra las estimaciones de plataformas web especializadas, confirmando la validez del modelo.

Palabras Clave: Tiempos de viaje, Regresión, Vecinos más cercanos, Uber Movement

Abstract. Traffic congestion is one of the most important factors affecting the quality of life in large cities. It increases travel times and, as a consequence, the urban distribution of goods. In distribution logistics, neglecting its effects may lead to the creation of inefficient routes, wasting resources, increasing pollution and decreasing vehicle load factors. This paper addresses the forecast and modeling of travel times depending on the time of day and the direction of movement between pairs of locations in the city of Bogotá. It is proposed to exploit the consolidated information, freely available, offered through the Uber Movement® website to construct models representing the variations on travel time observed during the course of a typical business day. K-nearest neighbors regression is used to predict the travel time required for travelling between pairs of customers. Next, the predictions are used to construct a model based on a piecewise linear

function representing the average time required for each moment throughout the day. The validity of the model is verified by comparing the time required for performing a tour using the estimations against those computed using specialized web platforms.

Key Words: Travel times, K-nearest neighbor regression, Uber Movement

INTRODUCCIÓN

La estimación y predicción de los tiempos de recorrido, considerando la congestión vehicular, juega un rol central en la toma de decisiones para diferentes actores en materia de planeación urbana (Mori et al. 2015). En la medida que las predicciones son confiables, los viajeros pueden planear sus recorridos y ahorrar tiempo de viaje, las entidades gubernamentales pueden incorporarla en la generación de políticas públicas o en actividades de planeación, entre otros. En logística, las predicciones de tiempos de trayecto pueden ayudar a reducir los costos de entrega, aumentar su confiabilidad, y mejorar la calidad del servicio. Para las compañías prestadoras de servicios logísticos (PSL), tal información es un insumo clave que podría contribuir a mejorar la planeación de actividades de transporte y distribución (Maden et al. 2010; Kok et al. 2012) y tener un impacto para aliviar los efectos de la congestión (Schmitt y Julia 2006).

Aunque es común encontrar herramientas para estimar el tiempo de recorrido entre puntos geográficos (e.g. Google maps), rara vez es posible acceder a esta información libremente para utilizarla a gran escala, por ejemplo, para la generación de rutas para la distribución de mercancías. En este trabajo exploramos la utilización de datos consolidados abiertos para generar modelos que se ajusten a las particularidades de la variación diaria de los tiempos de recorrido en entornos urbanos. Aun si hay información disponible, no necesariamente es completa. En particular, se propone combinar predicciones obtenidas a partir de datos históricos, con interpolaciones para datos faltantes, mediante funciones lineales por partes para estimar los tiempos de recorrido que puedan ser utilizados en, por ejemplo, la resolución del problema de ruteo de vehículos con tiempos de desplazamiento dependientes de la hora.

METODOLOGÍA

Modelando los tiempos de recorrido

A diferencia de algunos trabajos previos que se han focalizado en predecir los tiempos de pequeños segmentos de la malla vial (Mori et al. 2015), los cuales son agrupados para representar trayectos más largos, en este trabajo se propone utilizar los datos consolidados de múltiples opciones de desplazamiento entre dos locaciones (a veces reactivos a las condiciones). En particular se plantea utilizar la información disponible, obtenida de la base de datos Uber Movement ® en la generación de modelos del tiempo de desplazamiento de la ciudad de Bogotá. Dicha fuente provee, entre otras

Artículo No. 06. No. 10, Vol. V, Año 5, 2019. Revista de divulgación y tecnología de la Red Internacional de Investigadores de Ingeniería Industrial <REDI4>

métricas, los promedios históricos de los tiempos de recorrido entre algunos barrios de la ciudad en diferentes momentos del día. Por cuestiones de confidencialidad, la información solo se encuentra disponible cuando el número de viajes entre dichas localidades permite anonimizar la identidad de quienes aportan los datos. Este hecho implica que existen tiempos de trayecto entre pares de barrios (localidades) para los cuales no se cuenta con información. Por otro lado, los datos disponibles tienen muestras lo suficientemente grandes para anonimizar las identidades y, por lo tanto, son estimaciones robustas de los tiempos de trayecto.

Estimación de datos faltantes: Regresión mediante el método de los k vecinos más cercanos

Con el fin de estimar la duración de los recorridos cuando hay información faltante: la información no se encuentra en la base de datos de los tiempos de desplazamiento para la ciudad de Bogotá, se propone la utilización del método de los k-vecinos más cercanos (Fix & Hodges, 1951). El objetivo es determinar, a partir de la información disponible, la duración de los recorridos entre puntos faltantes. El propósito de la regresión es suavizar los datos disponibles para generar predicciones. El primer paso del método consiste en seleccionar un parámetro k , que representa el número de vecinos que conformaran el vecindario $\mathcal{N}$, i.e. $|\mathcal{N}| = k$, y que se consideraran en la obtención de la predicción. La selección de estos vecinos se hace considerando su “distancia” a los datos de entrada usando alguna métrica entre múltiples posibles (Distancia Manhattan, Euclidiana, de Hamming, etc.).

Sea y' el tiempo promedio de desplazamiento y X' un arreglo correspondiente a la longitud y latitud de los puntos de origen y destino y la hora del día para el cual se tiene el promedio. En su caso más simple, la predicción puede ser calculada como $y^* = \frac{1}{|\mathcal{N}|} \sum_{y' \in \mathcal{N}} y'$. Una versión alternativa a este proceso pondera cada uno de los vecinos usando pesos w' de forma tal que los vecinos más cercanos tengan más influencia sobre la predicción que aquellos más lejanos.

Interpolación de los tiempos de recorrido mediante función lineal por partes

La información generada tras utilizar el modelo de regresión mediante el método de los vecinos más cercanos permite determinar, de forma aproximada, los tiempos esperados ($\mu_{\#}$, μ_{+} , ... μ_{45}) para los desplazamientos que tomaran lugar en cada una de las horas del día. No obstante, estas predicciones no toman en cuenta las variaciones que ocurren al interior de cada uno de estos intervalos.

En este trabajo se propone modelar al cambio de los tiempos entre horas mediante la incorporación de la información disponible en la construcción de una función lineal por partes (ver Figura 1). Se asume que las variaciones en el comportamiento del tiempo de viaje se dan sobre la mitad de los intervalos analizados (i.e. 00:30, 01:30, ..., 23:30). La idea básica consiste en determinar el tiempo de desplazamiento en la mitad de cada intervalo de una hora ($\tau_{\#}$, τ_{+} , ... τ_{45}) se respeta los tiempos promedio de desplazamiento estimados o pronosticados.

El proceso se resume en utilizar las predicciones disponibles como información de entrada y hallar una solución a un sistema compuesto por $T=24$ ecuaciones lineales sobre T incógnitas.

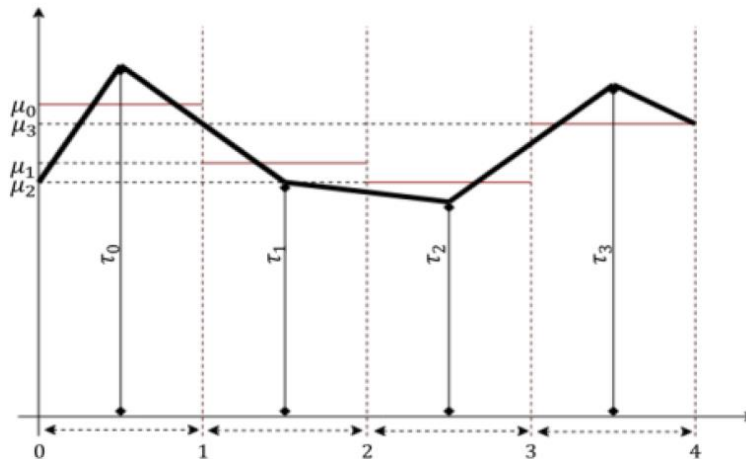


Figura 1. Parámetros de la función lineal por partes usada para modela los tiempos de desplazamiento

RESULTADOS

Con el propósito de evaluar los enfoques propuestos se propone utilizar como datos de entrenamiento los centroides de cada uno de los barrios para los cuales Uber Movement[®] ofrece información. Considerando que los tiempos de desplazamiento en el interior de cada uno de estos barrios está en el orden de 1 a 3 minutos, asumimos que calcular el tiempo entre pares de estos permite representar de forma apropiada los tiempos de desplazamiento para la ciudad. La selección de los parámetros del método (k , métricas de distancia entre vecinos, ponderación de vecinos) se realiza mediante una validación cruzada usando más de 8 millones de entradas de información disponible para el primer trimestre de 2018. Las predicciones se realizan utilizando la librería Scikit Learn sobre Python 3.6.5.

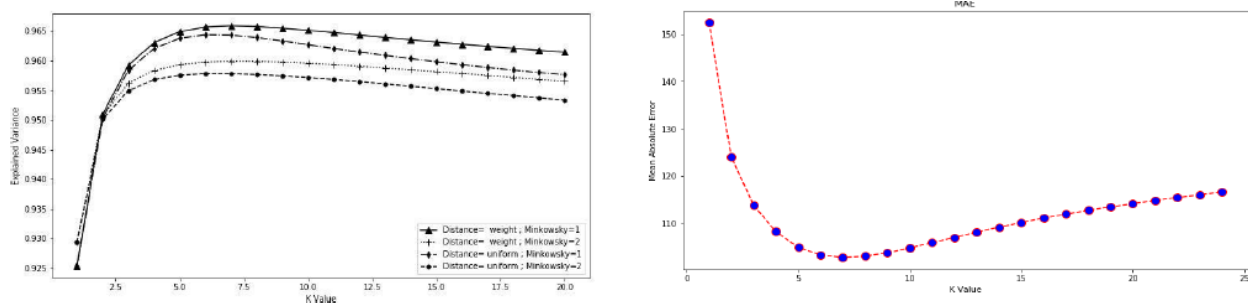


Figura 2. Varianza explicada para diferentes parámetros del método de los vecinos más cercanos (izq.) y Desviación media absoluta (MAE) para el modelo con los mejores resultados.

Artículo No. 06. No. 10, Vol. V, Año 5, 2019. Revista de divulgación y tecnología de la Red Internacional de Investigadores de Ingeniería Industrial <REDI4>

Los resultados de la validación cruzada (Figura 2) sugieren que las mejores estimaciones se obtienen cuando se ponderan los tiempos de desplazamiento de los 7 vecinos más cercanos utilizando una distancia Manhattan (Minkowski=1) y ponderando con mayor importancia los vecinos más cercanos (en este caso de acuerdo con el inverso de sus distancias). Bajo estas condiciones se alcanza una varianza explicada de 0.97 y una desviación media absoluta sobre los 100 segundos. La validación indica también que el 90% de las predicciones se desvían

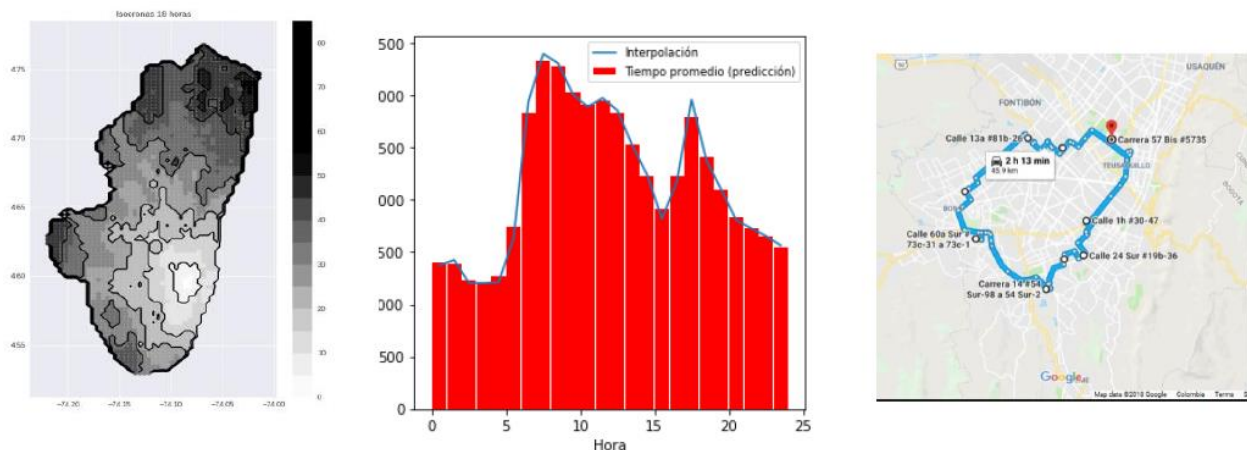


Figura 3. (Izquierda) Líneas Isócronas para los viajes saliendo de la plaza de Bolívar a las 18 horas (Centro) Tiempo de desplazamiento estimado entre los barrios cedritos y las Aguas (lunes a viernes). Estimación de una ruta usando el servicio de Google Maps® (Derecha)

Una última verificación de los resultados del enfoque se realiza comprobando el tiempo total requerido para ejecutar rutas concretas. Se utiliza la solución a un problema del agente viajero con tiempos de desplazamiento dependientes de la hora. Utilizando las predicciones del modelo se generaron recorridos de máximo 8 clientes (Figura 3). Los resultados sugieren que las predicciones del modelo propuesto representan apropiadamente los tiempos de desplazamiento cuando las rutas son comparadas contra los tiempos reportados por Google Maps ®.

Agradecimientos

Los autores agradecen a Colciencias la participación de Fabián Castaño en este proyecto, apoyada mediante la financiación otorgada a través del programa de apoyo a estancias postdoctorales Convocatoria 784-2017.

REFERENCIAS

Fix, E., & Hodges Jr, J. L. (1951). "Discriminatory analysis-nonparametric discrimination: consistency properties". *International Statistical Review/Revue Internationale de Statistique*, 57(3), 238-247.

Artículo No. 06. No. 10, Vol. V, Año 5, 2019. Revista de divulgación y tecnología de la Red Internacional de Investigadores de Ingeniería Industrial <REDI4>

Kok, A. L., et al. "Vehicle Routing under Time-Dependent Travel Times: The Impact of Congestion Avoidance." *Computers and Operations Research*, vol. 39, no. 5, Elsevier, 2012, pp. 910–18, doi:10.1016/j.cor.2011.05.027.

Lin, He, et al. "A Review of Travel-Time Prediction in Transport and Logistics." ... *Eastern Asia Society for Transportation ...*, vol. 5, no. March, 2005, pp. 1433–48, <http://front.cc.nctu.edu.tw/Richfiles/15767-1433.pdf>

Maden, W., et al. "Vehicle Routing and Scheduling with Time-Varying Data: A Case Study." *Journal of the Operational Research Society*, vol. 61, no. 3, Palgrave Macmillan, 2010, pp. 515–22, doi:10.1057/jors.2009.116.

Mori, Usue, et al. "A Review of Travel Time Estimation and Forecasting for Advanced Traveller Information Systems." *Transportmetrica A: Transport Science*, vol. 11, no. 2, 2015, pp. 119–57