

Tiempos de traslado y de servicio en la distribución de mercadería de última milla. Estimación mediante técnicas de minería de datos

Emiliano Bodean¹[0009-0001-4902-7605] y Gabriela P. Henning^{2, 3}[0000-0002-2660-0376]

¹ Facultad Regional Paraná, UTN, Almafuerde 1033, 3100 Paraná, Argentina

² INTEC (UNL, CONICET), Güemes 3450, 3000 Santa Fe, Argentina.

³ Facultad de Ingeniería Química, UNL, Santiago del Estero 2829, 3000 Santa Fe, Argentina
emilianobodean@frp.utn.edu.ar - ghenning@intec.unl.edu.ar

Abstract. Se presentan avances en el desarrollo de una metodología de estimación de tiempos de traslado y de servicio en los problemas de distribución de última milla de mercadería, empleando herramientas de minería de datos. Estos tiempos, que son afectados por una multiplicidad de factores, son utilizados como datos de entrada en problemas de tipo VRPTW (*Vehicle Routing Problem with Time Windows*) que se presentan en áreas urbanas. La correcta estimación de dichos tiempos juega un rol central en la efectiva solución de estos problemas y la consecuente eficiencia de los sistemas de distribución gestionados. Asimismo, resultan cruciales en los sistemas de transporte inteligentes asociados a las *Smart Cities*. Para realizar las estimaciones ha sido necesario anonimizar, limpiar, pre-procesar y analizar una importante cantidad de datos crudos provenientes de las bases de una empresa de seguimiento satelital de vehículos. Este trabajo describe dichas actividades y sus resultados, al igual que las tareas que se están llevando a cabo en la actualidad en relación a la incorporación de datos de contexto a los *dataset* de trabajo. También se describen las actividades futuras vinculadas a la solución del problema de estimación propiamente dicho.

Keywords: Redes Urbanas, Tiempos Traslado, Tiempos Servicio, Última Milla.

1 Introducción

Un eslabón muy importante de las Cadenas de Suministros asociadas a productos de consumo es el transporte de última milla, también conocido como distribución capilar o entrega final, el cual hace referencia al transporte urbano de mercaderías, es decir, a la etapa final de la cadena de distribución [1]. La distribución física de mercaderías abarca un conjunto variado de problemas, entre los que se destacan los de ruteo de vehículos (VRP: *Vehicle Routing Problem*) [2]. Algunos de los aspectos que influyen en la distribución urbana de mercaderías son los plazos de entregas reducidos, no sólo los relacionados a la

inmediatez del envío luego de realizada la compra, sino también los vinculados a las ventanas horarias estrechas impuestas por los clientes. Esta problemática se identifica como VRPTW (*Vehicle Routing Problem with Time Windows*) [1-3]. Entre los datos básicos de estos problemas se encuentran los tiempos de traslado y de servicios. En un entorno urbano con congestión, las velocidades no son constantes y el tiempo de traslado entre dos puntos, por lo general, no es sólo función de la distancia recorrida. La densidad del tráfico varía y causa fluctuaciones en la velocidad de desplazamiento y en los tiempos de traslado [4]. Por esto, la estimación de los tiempos del traslado dependerá de diversos factores de la red vial, especialmente en las zonas urbanas, tales como el día de la semana, la hora y las condiciones del tránsito, así como las condiciones meteorológicas. Pero también será función del tipo de vehículo, el tipo de carga, el chofer, entre otros. Por otro lado, los tiempos de servicio asociados a los puntos visitados también son parámetros que poseen variabilidad y dependen de diferentes factores como las características del nodo, las particularidades de la carga, la ubicación de la misma en el vehículo, entre otros. Si los tiempos de traslado y de servicio utilizados como datos de entrada en problemas de tipo VRPTW se apartan de los tiempos reales, se obtendrán soluciones alejadas de la realidad, con demoras en las entregas, e inclusive incumplimiento de los horarios pautados. Por otra parte, contar con buenas estimaciones de los tiempos de traslado entre puntos de una red urbana juega un rol central en los sistemas de transporte inteligentes asociados a las actuales *Smart Cities*.

Se presentan avances en el desarrollo de una metodología de estimación de tiempos de traslado y de servicio en problemas asociados a la distribución de última milla mediante herramientas de minería de datos. Ésta es un área de sumo interés, tal como lo revela la síntesis del análisis bibliográfico efectuado (Sección 2). Para realizar las estimaciones ha sido necesario anonimizar, limpiar, pre-procesar y analizar una importante cantidad de datos crudos provenientes de las bases de una empresa de seguimiento satelital de vehículos, ubicada en la Ciudad Autónoma de Buenos Aires (CABA), que para cada recorrido urbano incluyen: vehículo utilizado, múltiples datos del dispositivo GPS instalado en el camión/camioneta, clientes visitados, tipo de carga transportada, chofer, entre otros. Los resultados parciales de esta labor se exponen en la Sección 3. Finalmente, en la Sección 4 se describen las tareas que se están llevando a cabo en la actualidad y las actividades futuras.

2 Análisis bibliográfico

En los últimos años se han desarrollado diversos métodos de estimación de tiempos de traslado, los cuales se basan en diferentes metodologías y tienen sus pros y contras, como se analiza en [5]. Esta revisión describe las propuestas existentes y las clasifica en métodos basados en modelos y basados en datos. También compara las metodologías en términos de conjuntos de datos, rango de predicción y precisión. Con la llegada del Internet de las cosas y las ciudades inteligentes, hay más datos disponibles, especialmente en entornos urbanos, y los métodos basados en datos han ido ganando cada vez más relevancia. Por ello, y por la

naturaleza del problema abordado, el análisis bibliográfico se centró en este tipo de propuestas. Por falta de espacio, sólo se señalan algunos trabajos relevantes [6 -7].

3 Resultados Parciales Obtenidos

La obtención de la información relevante se inició con el tratamiento de los datos contenidos en tres bases distintas de la empresa: itinerarios, posiciones y vehículos. La primera comprende el historial de recorridos planificados para la flota de vehículos durante un período de tiempo. Cada itinerario incluye datos sobre: origen, secuencia de nodos a visitar, vehículo, fecha, entre otros. La base de posiciones contiene datos de todos los recorridos efectivamente realizados por los transportes, los cuales fueron reportados por los dispositivos GPS instalados en los vehículos. Por su parte, la base de vehículos presenta datos (marca, modelo y otras características relevantes) de todos los vehículos involucrados.

El administrador de base de datos utilizado durante el trabajo es PostgreSQL14. Además, se implementaron scripts específicos en lenguaje R para efectuar el filtrado, la descarga y anonimización de los datos relevantes, los cuales fueron almacenados en formato CSV. A partir de esta información, y con un adecuado procesamiento, se obtuvieron nuevos conjuntos de datos, el *dataset* de tiempos de trasladado entre pares de puntos de servicio y el de tiempos de servicio. Cabe señalar que en forma previa se llevó un análisis exploratorio de las bases de datos existentes para conocer mejor los datos, identificar las variables relevantes y facilitar los procedimientos de limpieza que se llevaron a cabo. Asimismo, se identificaron factores críticos que afectan la precisión de las mediciones de los tiempos de traslado en base a los datos brindados por los dispositivos de seguimiento vehicular, a saber: **Método de adquisición de datos de ubicación del vehículo:** Los dispositivos GPS instalados en los vehículos operan con una lógica de adquisición de datos parametrizada en función del tiempo, la distancia, la dirección y la velocidad. Estos parámetros se ajustan típicamente para garantizar que los puntos registrados en el mapa proporcionen una representación adecuada de las rutas, evitando la generación de datos redundantes. La precisión de la medición de los tiempos de arribo y partida depende en gran medida de la forma en que se adquieren estos datos. Por ejemplo, registrar la posición del vehículo cuando su velocidad es cero permite identificar el momento exacto de la llegada.

Definición precisa de las geocercas de los nodos a visitar: En entornos urbanos, es común que la entrega de mercancías se realice en la vía pública, lo que implica que la descarga puede ocurrir en las inmediaciones del nodo de destino. Es crucial definir geocercas adecuadas que abarquen un área suficientemente amplia para detectar con precisión la llegada a estos puntos, pero tampoco excesivamente extensas. Se encontraron casos donde los nodos estaban tan próximos entre sí que resultaba imposible diferenciar los tiempos de llegada a cada uno de ellos y hubo que dar un tratamiento especial a estas situaciones.

Estas consideraciones son fundamentales para garantizar la fiabilidad y la precisión de la medición de los tiempos de traslado en el contexto del seguimiento vehicular.

4 Actividades en curso y tareas futuras

A los datos descritos en la sección previa se les pretende agregar nuevas variables. Se están evaluando: (i) el clima, incorporando variables de temperaturas y alertas meteorológicas de cada día, mediante el Catálogo de Datos Abiertos del Servicio Meteorológico Nacional, (ii) el calendario de días feriados y (iii) datos de flujo vehicular por día en diferentes puntos de CABA. Asimismo, se evaluarán los posibles beneficios de la discretización de los datos espaciales mediante el uso de *Geohash* u otro sistema similar. *Geohash* es un sistema de codificación geográfica que cataloga una ubicación usando una cadena corta de letras y dígitos. Se trata de una estructura jerárquica de datos espaciales que subdivide el espacio en cubos de forma de cuadrícula [8]. Esta propuesta surge pues las ubicaciones continuas con alta precisión no permiten encontrar patrones comunes en diferentes muestras.

Se continuará avanzando con el desarrollo de modelos predictivos. Se comenzaron a utilizar los algoritmos basados en *Gradient Boosting* de la librería XGBoost de R [9]. En la generación de los modelos, se realizará la optimización de hiperparámetros. Algunas de las opciones disponibles para esta tarea son paquetes basados en algoritmos genéticos, en optimización bayesiana u otros. Estos modelos se validarán con técnicas que separan los datos en conjuntos de entrenamiento y prueba como, por ejemplo, la técnica de validación cruzada.

Referencias

1. Andino, E.: ¿Qué es la última milla? Énfasis Logística, XXVII(4), 46-48. <https://logisticasud.enfasis.com/revista/?embtrk=5a2h-R-44320496-R-7g1fj7-R-5h2dk5>: (2021)
2. Paraskevopoulos, D. C., Laporte, G., Repoussis, P. P., Tarantilis, C. D.: Resource constrained routing and scheduling: Review and research prospects. *European Journal of Operational Research*, 263(3), 737-754 (2017)
3. Braekers, K., Ramaekers, K., Van Nieuwenhuyse, I.: The vehicle routing problem: State of the art classification and review. *Computers & Industrial Engineering*, 99, 300-313 (2016)
4. Miguel, F., Frutos, M., Tohmé, F.: Programación de Operaciones Logísticas de Distribución Urbana de Mercadería. Primera Edición. Editorial de la Universidad Nacional del Sur (2017)
5. Bai, M., Lin, Y., Ma, M., Wang, P.: Travel-Time Prediction Methods: A Review. In: Qiu, M. (eds) *Smart Computing and Communication. SmartCom 2018. Lecture Notes in Computer Science*, vol 11344. Springer, Cham (2018)
6. Zhang, Y., Haghani, A.: A gradient boosting method to improve travel time prediction. *Transp. Res. Part C* 58, 308–324 (2015).
7. Liu, Y., Wang, Y., Yang, X., Zhang, L.: Short-term travel time prediction by deep learning: a comparison of different LSTM-DNN models. In: *IEEE International Conference on Intelligent Transportation Systems*, pp. 1–8 (2017)
8. Yus, D.: Long-term vehicle movement prediction using Machine Learning methods. <http://www.diva-portal.org/smash/get/diva2:1241645/FULLTEXT01.pdf> (2018)
9. XGBoost. <https://cran.r-project.org/web/packages/xgboost/index.html> (2024)