

# **Automatización en la Captura de Datos para el Modelado de Flujo Vehicular**

Julio Monetti, Mariana Brachetta y Oscar León  
Universidad Tecnológica Nacional – Facultad Regional Mendoza  
Rodríguez 273- Mendoza - Argentina  
{jmonetti,mbrachetta,oleon}@frm.utn.edu.ar

**Resumen.** El trabajo presenta experiencias en el modelado de datos sobre la dinámica vehicular en zonas urbanas. A partir de la investigación sobre software que permita realizar modelos de tránsito, surge luego la propuesta de crear nuevas estructuras de datos, modelos para simulación y un software que permita adecuarse a las necesidades particulares de investigación. Se pretende a través del presente estudio proponer metodologías para la captura y procesamiento de datos, no solo proveyendo formato sobre los datos obtenidos del modelado y posteriores simulaciones, sino que el aporte se realice a la luz de una experiencia de la computación aplicada al análisis y diseño de productos y datos útiles para la ingeniería de tránsito. La aplicación de sistemas GPS proporciona la automatización requerida para la obtención de datos masiva. Complementa el trabajo el modelado de datos para un adecuado almacenamiento.

**Palabras Clave:** Flujo Vehicular, Modelos de Tránsito, Congestión Vehicular.

## **1 Introducción**

El proyecto en el cual se encuadra el trabajo tiene como objeto describir las características de la dinámica del flujo vehicular en ambientes urbanos, para la elaboración de una metodología estándar que automatice la captura, procesamiento y almacenamiento de datos, identificando las variables que determinan situaciones particulares, como por ejemplo la congestión vehicular [1]. Se busca reemplazar metodologías clásicas de aforo vehicular, donde generalmente se utilizan hojas de datos para asentar mediciones sobre velocidad, densidad, etc. La utilización de técnicas alternativas, como por ejemplo la del auto flotante [2], representa un esfuerzo para explicar inestabilidades del tránsito en función de la respuesta del conductor. Al utilizar un vehículo testigo circulando por las áreas de estudio, es posible obtener mediciones directamente desde el mismo. Tales mediciones

corresponden a la obtención de velocidades puntuales, que son utilizadas en etapas posteriores para establecer condiciones de congestión vehicular.

En la sección 2 se mencionan estudios preliminares que se encuadran en el tema de estudio. La sección 3 describe el diseño físico de la arquitectura computacional utilizada para el almacenamiento y cómputo de los datos recolectados: sus bases de datos, y una descripción de los algoritmos utilizados. En la sección 4 se menciona el modelo de datos propuesto, donde las características estructurales de las tablas de datos resultan de suma importancia para mantener un rápido acceso a los mismos. La sección 5 hace referencia a uno de los productos considerados por el grupo de trabajo: algoritmos para establecer rutas más cortas. Finalmente se presentan las conclusiones sobre el análisis, diseño, desarrollo y uso del sistema de información planteado.

## **2 Estado del Arte**

### **2.1 Productos para la Asistencia al Modelado de Tránsito**

Los estudios sobre tránsito y transporte [3][4], requieren del modelado y simulación de escenarios de circulación; lo cual requiere una correcta determinación de aquellos indicadores que permitan describir en forma clara y precisa situaciones y proyecciones de la dinámica vehicular. Se apunta a ofrecer soluciones alternativas sobre el procesamiento de datos provenientes de dispositivos GPS, atendiendo a la estructura de datos y algoritmos considerados para el planteo de escenarios particulares de flujo vehicular, determinación de indicadores de congestión, etc. Se indagó sobre el mercado de software destinado al modelado vehicular [5], encontrando que existen variadas metodologías para muestreo vehicular (por ejemplo, el método del auto flotante entre otros), y las correspondientes herramientas para simulación de escenarios de tránsito. Se han analizado distintas perspectivas que presentan estos productos para el análisis a través de la generación de micro y macro modelos. Dentro de las herramientas que se encuentran disponibles actualmente en el mercado, se destaca el software AIMSUN [6], el cual ha sido tomado como modelo debido al grado de generalidad que presenta. Las capacidades provistas por el software se centran principalmente en la realización de micro y macro modelos de una red de tránsito, como así también la simulación de maniobras reales. No obstante ello, el software no ha permitido avanzar sobre puntos de estudio u observación claves dentro de situaciones planteadas, por ejemplo: determinación de la congestión vehicular a través de datos históricos. Este estudio resulta sumamente conveniente para establecer políticas de circulación o diseño de obras civiles.

### **2.2 Estudios de Campo**

En primer lugar se han utilizado hojas de datos para la recolección de datos sobre velocidades y densidad vehicular, los cuales permiten en una primera instancia analizar las posibilidades de automatización de las muestras, el análisis de

dispositivos de captura, etc. La experiencia resulta de utilidad para observar tanto las principales variables de medición (por ejemplo: velocidad media por tramo), como así también situaciones de circulación irregulares o anómalas. A partir de los estudios mencionados, se decide la implementación de herramientas propias para el almacenamiento, procesamiento y generación de información sobre los diferentes escenarios modelados.

### 3 Arquitectura Computacional de la Solución Propuesta

#### 3.1 El dispositivo de Captura de Datos

Para la obtención de trazas se utiliza un dispositivo *GPS GARMIN Etrex Vista HCx* [7], el cual resulta adecuado para obtener datos que se ajusten tanto a las variables de medición requeridas para el posterior modelado, como así también al tiempo mínimo requerido entre registros (por ejemplo: cada un segundo). Por otro lado se analizó la aplicación de *SmartPhones* para la adquisición masiva de datos, ya que por su popularidad, es posible distribuir la toma de datos a lo largo de mayor cantidad de vehículos de prueba. Se considera por otro lado que estos dispositivos son extremadamente dependientes del software que poseen, no encontrando estándares que permitan una obtención de datos homogéneamente confiables en el corto plazo.

#### 3.2 Diseño de la Aplicación

Si bien el sistema de información planteado no cuenta con las características de un sistema de procesamiento en tiempo de real, se pretende automatizar la captura del tren de datos provenientes del dispositivo GPS, de tal forma que resulte conveniente para la generación de información estadística. La figura 1 muestra en forma esquemática los diferentes subsistemas considerados para la obtención de información estadística.



**Fig. 1.** El sistema de información propuesto tiene por objeto el almacenamiento, procesamiento de datos y la posterior generación estadística sobre los escenarios modelados.

En primer lugar se cuenta con un conjunto de funciones que permiten el reconocimiento de diferentes tipos de trazas (*Parser*). Luego, a partir de la depuración de las trazas capturadas por los dispositivos, los datos son filtrados, eliminando aquellas muestras que no son útiles para el posterior procesamiento

(*Filtros y Conversion*). A continuación se trabaja sobre algoritmos que procesan los datos filtrados para ser almacenados en base de datos. Finalmente, se requiere del procesamiento de la información para la generación de escenarios posibles de circulación. Esto plantea el problema de tener que procesar grandes volúmenes de datos. El producto obtenido consiste en un conjunto de algoritmos, bases de datos normalizada y metodologías de procesamiento, que sirve como insumo para ser utilizado en futuros proyectos de modelado numérico de situaciones reales de tránsito. En resumen: el aporte esperado es la confección de una metodología de procesamiento masivo de datos de la dinámica vehicular, lo cual permita obtener información útil para crear nuevos modelos urbanos. Un avance del trabajo realizado se menciona en la sección 4.

### 3.3 Producción de la Información

La capa de *Producción* (figura 1) corresponde a un conjunto de algoritmos que establecen *a priori* información útil para la circulación de un vehículo sobre el área de estudio. Los algoritmos se orientan a la sumarización, ordenamiento, agrupamiento, etc, de conjuntos de datos provenientes de las trazas, y de acuerdo a metodologías estándares de muestreo establecidas por la Ingeniería de Tránsito. Las muestras correspondientes a los conjuntos de datos en bruto, provenientes de numerosas muestras obtenidas a través de los dispositivos de georeferenciación, son sistematizadas a fin de generar información para elaborar escenarios de dinámica vehicular. El procesamiento principal corresponde a la obtención de caminos más cortos entre dos puntos específicos del área de estudio, contando con información estadística sobre la circulación del vehículo de prueba sobre dicha área.

### 3.4 Almacenamiento y Clasificación

El almacenamiento principal se encuentra en una primera instancia sobre un conjunto de archivos de texto obtenidos del dispositivo. Se ha propuesto una metodología estándar de clasificación de tales ficheros para mantener un orden de acuerdo al vehículo testigo que toma la muestra, y otras variables que permiten describir el entorno en el cual se tomo la muestra, por ej: Característica del Vehículo Testigo, Día de la Semana / Fecha, Hora, Datos climáticos, etc.

Cada fichero representa una circulación particular del vehículo de prueba; luego, el conjunto de ficheros es introducido en la base de datos con el objeto de generar un único dominio de datos del cual poder generar información estadística. Dado la necesidad de almacenamiento masiva, y con el objeto de mantener almacenadas la totalidad de las muestras obtenidas del conteo vehicular, se requiere el desarrollo de un modelo de datos eficiente, para ser implementado con un motor de base de datos que resulte adecuado para la gestión de los datos. La confección de la base de datos consta de dos etapas:

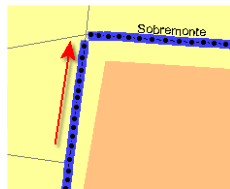
1. **Información Estática.** Comprende la georeferenciación de puntos de interés a lo largo de las zonas de estudio. Estos puntos luego conforman una malla

interconectada. La información estática está conformada por tablas que contienen básicamente la información sobre esquinas, tramos de calles (por ejemplo, en la figura 2 se muestra un tramo en azul, cuyos datos representan las coordenadas entre las cuales está comprendido, su nombre, etc.) y zonas de estudio; sus nombres y datos principales para identificarlas.



**Fig. 2.** La base de datos estática contiene información sobre coordenadas de puntos de interés e información sobre tramos.

- 2. Información Dinámica.** Corresponde a la información obtenida a través de la captura permanente de datos. Está compuesta por velocidades promedio, máximas, mínimas por tramo, conjunto de tramos etc. La contrastación de esta información con la malla obtenida en la etapa 1 permite generar información estadística sobre la dinámica vehicular, observar zonas de congestión, etc.



**Fig. 3.** La base de datos dinámica contiene información sobre recorridos de los vehículos testigos (velocidades y características de circulación).

Como se especifica en la sección 4, los datos almacenados en la base de datos son útiles para generar en tiempo de ejecución información útil para representar diferentes escenarios de circulación, por ejemplo a través de la constitución de matrices de adyacencia que expongan el paso entre diferentes puntos de la malla. El recorrido continuo del vehículo por la zona de prueba (ver figura 3) permite establecer el sentido de circulación entre diferentes tramos.



(a)

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(b)

**Fig. 4.** El sentido de circulación de cada tramo permite establecer un grafo dirigido (a), y a continuación una matriz de adyacencia que lo contenga (b).

Para el ejemplo de la figura 4 se han tomado solamente 6 tramos de estudio (fig. 4a), lo que deriva en una matriz de adyacencia (fig. 4b) de 36 coeficientes con solamente 5 elementos no nulos. Esta información es generada por el vehículo testigo en forma dinámica a medida que circula a través de los puntos (esquinas) almacenados previamente en la base de datos estática.

### 3.5 Algoritmos de Filtrado de Datos

Se desarrollan algoritmos para ajustar los datos de acuerdo a las necesidades de información, como por ejemplo la existencia de situaciones de circulación vehicular anómalas, fallas del dispositivo de captura de datos, etc. Por ejemplo, en una captura de datos es común encontrar numerosas muestras con  $vel=0$  (velocidad igual a cero, o auto detenido), o la captura de datos donde el vehículo testigo realiza diferentes maniobras. Estos eventos resultan inútiles para establecer información de interés, como por ejemplo la velocidad de marcha.

### 3.6 Reconocimiento de trazas heterogéneas

Estos algoritmos están destinados a la conversión de unidades (por ej. coordenadas con diferente formato) y eliminación de datos inútiles. A partir de los diferentes formatos obtenidos, resulta útil establecer una metodología de reconocimiento. Para ello se genera un compilador capaz de reconocer diferentes formatos de traza, proveniente de diferentes dispositivos. La figura 5 ejemplifica un formato básico de traza donde se observan entre otros datos, la coordenada y la velocidad puntual.

...			
75	05/07/2013 18:30:49	0 m 0:00:01	0 km/h S32.89669 W68.85365
76	05/07/2013 18:30:50	0 m 0:00:01	0 km/h S32.89669 W68.85365
77	05/07/2013 18:30:51	8 m 0:00:01	29 km/h S32.89669 W68.85365
78	05/07/2013 18:30:52	1 m 0:00:01	5 km/h S32.89662 W68.85364
79	05/07/2013 18:30:53	2 m 0:00:01	6 km/h S32.89661 W68.85364
80	05/07/2013 18:30:54	2 m 0:00:01	6 km/h S32.89659 W68.85364
81	05/07/2013 18:30:55	2 m 0:00:01	8 km/h S32.89658 W68.85363
82	05/07/2013 18:30:56	3 m 0:00:01	10 km/h S32.89656 W68.85363
83	05/07/2013 18:30:57	3 m 0:00:01	12 km/h S32.89653 W68.85363
84	05/07/2013 18:30:58	3 m 0:00:01	12 km/h S32.89650 W68.85362
85	05/07/2013 18:30:59	3 m 0:00:01	11 km/h S32.89647 W68.85362
86	05/07/2013 18:31:00	3 m 0:00:01	10 km/h S32.89645 W68.85361
87	05/07/2013 18:31:01	3 m 0:00:01	10 km/h S32.89642 W68.85361
...			

**Fig. 5.** Ejemplo de una traza de datos. Cada línea representa un registro obtenido por el dispositivo de captura de datos.

El compilador de trazas [8], es desarrollado a través de las herramientas Flex y Bison [9] [10], y comprende las siguientes tareas: 1) Incorporación y lectura del fichero de trazas, 2) Determinación de errores en las mediciones y 3) Compilación de la traza hacia un formato estandarizado.

La consolidación de datos a partir de trazas heterogéneas permite luego obtener una mayor cantidad de muestras, contando con una base de datos con trazas provenientes de diferentes dispositivos o diferentes modalidades de captura.

## 4 Modelo de Datos

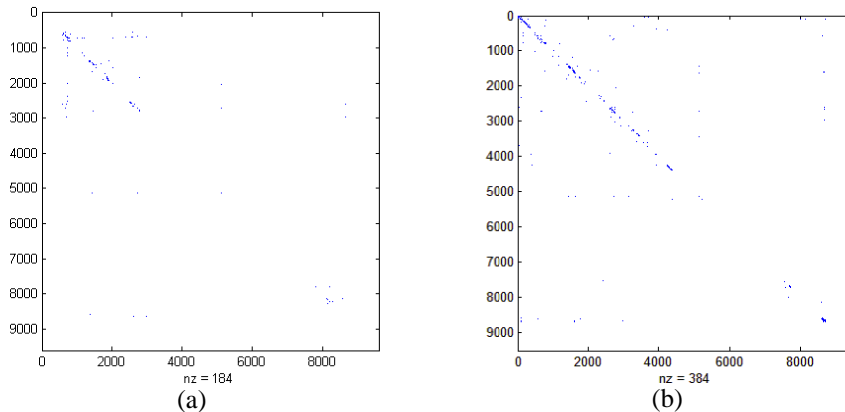
### 4.1 Adquisición de Datos para la base de Datos Estática

Se han realizado algoritmos que permiten a través de los datos obtenidos en crudo, establecer la existencia de esquinas (o puntos de interés), con lo cual es posible alimentar la base de datos en forma continua, y tras la circulación permanente del vehículo de prueba por zonas de estudio.

### 4.2 Adquisición de Datos para la base de Datos Dinámica

La base de datos con información dinámica contiene información recolectada a través de la técnica del auto flotante. La información obtenida se puede caracterizar en base a: 1) Información sobre el paso de un tramo a otro, 2) Información sobre la velocidad puntual en coordenadas específicas.

Se obtienen permanentemente datos a través del dispositivo GPS, los cuales alimentan la base de datos de *esquinas* y *tramos*. En la figura 6 se observa que a medida que se obtienen más trazas a través de la circulación libre del vehículo a través de la ciudad, se registran más datos *paso(tramo1,tramo2)*. A simple vista se observa mayor densidad en la matriz de adyacencia.



**Fig. 6.** La Matriz de Adyacencia presenta coeficientes que determinan el paso de un tramo *paso(fila,columna)*. (a) y (b) son dos momentos diferentes del estado de la base de datos, donde la segunda matriz presenta mayor cantidad de entradas.

Se puede visualizar en la figura que la densidad de la matriz se centra a lo largo de la diagonal principal, situación ventajosa al trabajar la matriz numéricamente.

### 4.3 Formato de Representación de la matriz de adyacencia

A partir del modelado de datos se cuenta con una matriz binaria de 81.000.000 coeficientes, donde una entrada diferente de 0 en  $(fila,columna)$  determina el paso entre  $(tramox,tramoy)$  donde  $tramox=fila$  y  $tramoy=columna$  del coeficiente. Esta metodología permite describir un grafo dirigido, donde los nodos representan a cada tramo de estudio, y las aristas el sentido de circulación vehicular entre dichos tramos.

Al almacenar la matriz completa en memoria RAM se requieren  $9.000*9.000*2 = 154,4952Mb$  (para un coeficiente que ocupa 2 bytes en memoria), siendo el porcentaje ocupación del  $4,74e^{-4}\%$ .

Al operar con la matriz de adyacencia (como se especifica en la siguiente sección) se torna necesario disponer de arreglos auxiliares, con lo cual, las magnitudes expresadas crecen considerablemente.

Se debe diferenciar el almacenamiento temporal de la matriz en memoria RAM, con la representación de la misma en disco. En este último se almacena la matriz completa con el objeto de mantener una mejor reutilización de datos entre los programas que utilizan tales ficheros. Por otro lado, se busca una representación adecuada de la matriz en memoria RAM, teniendo en cuenta las características de dispersión de la misma. La representación de tipo *coordenada*, como un tipo básico para la representación de una matriz rala [11] lleva a almacenar únicamente los elementos no nulos junto con datos que determinan su ubicación  $(fila,columna)$ . Este almacenamiento permite reducir la cantidad de espacio en memoria principal a menos de 3Kb para la matriz del ejemplo. Por otro lado, la complejidad algorítmica de los métodos numéricos que operan sobre ella aumenta al requerir el mapeo de un coeficiente sobre una nueva estructura de datos diferente a un arreglo bidimensional.



## 5 Algoritmos para Encontrar Trayectorias.

### 5.1 Algoritmos para encontrar rutas

Muchos estudios dentro de las ciencias de la computación hacen referencia al esquema de computación planteado, conociéndose en la literatura como el problema APSP (*All Pairs Shortest Path*) aquel que tiende a encontrar caminos más cortos (o favorables) entre dos puntos en un grafo. A partir de la matriz de adyacencia es factible plantear la solución a problemas de este tipo. En este trabajo se destaca una situación especial al presentarse una matriz de adyacencia de grandes dimensiones. Luego, dicho problema, que en la comunidad de ciencias de la computación presenta un constante desafío de resolución al intentar minimizar tiempos de resolución, se incrementa al contar con un grafo dirigido con una cantidad de componentes considerable.

Se propone en primer lugar una solución que lista y aísla desde la base de datos aquellas relaciones que permiten establecer el camino entre dos puntos particulares en la zona de estudio, podando el resto del grafo dirigido.

La potenciación de la matriz de adyacencia representa algunos de los cálculos más complejos desde el punto de vista computacional al consumir considerables recursos y tiempo de ejecución. Este cómputo es realizado una considerable cantidad de veces al describir el camino más ventajoso entre dos puntos. Se realiza la potenciación de la matriz como una sucesión de productos, donde

$$P=A^2=A*A. \quad (1)$$

representa la matriz de adyacencia para pasos de longitud 2 entre dos puntos de interés. Se recuerda que la complejidad algorítmica de dicho producto es  $O(n^3)$  para  $n$  cantidad de coeficientes. De acuerdo a la cantidad de coeficientes, para un almacenamiento convencional como el descrito en la sección 4.3 se prevé que la carga computacional es considerablemente alta para el cálculo de (1).

### 5.2 Procesamiento Paralelo de la Información

Con el objeto de paliar los altos tiempos de cómputo mencionados en 5.1, se propone la paralelización del proceso principal. Se procede a la paralelización del algoritmo de multiplicación sucesiva de matrices y se proponen dos algoritmos de multiplicación. En primer lugar, un algoritmo basado en memoria distribuida y paso de mensajes entre nodos para ser ejecutado en un *cluster* de computadoras. Esta solución se materializa a través de la utilización de la librería MPI [12].

Luego se propone un algoritmo con memoria compartida para ser ejecutado en una arquitectura superescalar multinúcleo, a través de la paralelización de hilos *Posix* [13].

En ambos casos se consigue obtener un *speedup* considerable frente a la solución secuencial.

## Conclusiones y Trabajo Futuro

El trabajo permitió establecer un sistema de información que abarca desde la captura de datos, hasta la generación de estadísticas en torno a información real de circulación de vehículos testigo. Dicha información revela situaciones particulares, que son difíciles de obtener a través del aforo vehicular clásico.

Se facilita a investigadores e ingenieros relacionados con el tránsito y transporte el modelado de escenarios de flujo vehicular, proveyendo una estructura y conjunto de datos de fácil almacenamiento, acceso y manipulación; posibilitando a los mismos acceder a datos procesados o en crudo: condiciones no prevista en la mayoría de los software de modelado de tráfico vehicular de mayor incidencia en el mercado.

Para aquellos procesos que demandan mayor cantidad de tiempo de cálculo se proveyó de algoritmos paralelos, con los cuales se consiguieron considerables mejoras en los tiempos de ejecución.

Como trabajo futuro, el grupo analizará tecnologías para el procesamiento en tiempo real del tren de datos como resulta el GPRS. También se prevé como trabajo futuro la cuantificación y documentación de la ganancia en tiempo de ejecución experimentada con las nuevas arquitecturas de procesamiento paralela.

## Referencias

1. Thompson I., Bull, A.: La Congestión del Tránsito Urbano: Causas y Consecuencias Económicas y Sociales. CEPAL (Naciones Unidas, División de Recursos Naturales e Infraestructura, Unidad de Transporte). Serie Recursos Naturales e Infraestructura (25). Chile, (2001). ISBN: 92-1-321865-6
2. Hall R.W.: Handbook of Transportation Science. 2ª Edición. Edited by Randolph W. Hall, University of Southern California. Kluwer's International Series. USA, (2003). ISBN: 1-4020-7246-5.
3. Cal, R., Reyes, M.: Ingeniería de Tránsito. Fundamentos y Aplicaciones 7ª Edición. Alfaomega. (1995). ISBN: 970-15-0109-8.
4. Schöbel, A.: Optimization in Public Transportation: Stop Location, Delay Management and Tariff Design in a Public Transportation Network. Georg-August University (Göttingen-Alemania). Springer. USA, (2006). ISBN: 978-0-387-32896-6.
5. Software para el Modelado de Tránsito. Julio Monetti. Disponible en <http://grid2.frm.utn.edu.ar/proyectotransito>.
6. TSS. AIMSUN. Disponible en [http://www.mctsoft.com/html/fr\\_mct.htm](http://www.mctsoft.com/html/fr_mct.htm).
7. GARMIN. <https://buy.garmin.com/en-GB/GB/prod8703.html>
8. Compilador de Ficheros de Trazas. J. Monetti. <http://grid2.frm.utn.edu.ar/proyecto>.
9. Aaby, A.: Compiler Construction using Flex and Bison.. Walla Walla College. (2005). Disponible en [aabyan@wwc.edu](mailto:aabyan@wwc.edu).
10. Levine, J.: flex & bison 1ª Edición. O'Reilly Media. USA, (2009). ISBN: 978-0596155971
11. Tewarson R.: Sparse Matrices.. Department of Applied Mathematics and Statistics. Academic Press Inc. USA, (1973). ISBN: 9780126856507.
12. Gropp, W., Lusk E. Thakur, R.: Using MPI-2. Advanced Features of the Message-Passing Interface. The MIT Press. Cambridge – Massachusetts. London, England. (1999). IBN: 0-262-057133-1.
13. Butenhof, D.: Programming with Posix Threads. Addison Wesley Longman Inc. USA, (1997). ISBN: 0-201-63392-2.