



# BIG DATA CHILE

TRANSANTIAGO COMO FUENTE DE DATOS. LOS DATOS PASIVOS PUEDEN AYUDARNOS A HACER UNA MEJOR GESTIÓN DE LA CIUDAD | Marcela Munizaga

BIG DATA ¿LA MISMA CERVEZA PERO CON OTRO ENVASE? | Juan Velásquez

LA NUEVA ERA DE DATOS EN ASTRONOMÍA | Faviola Molina

MANEJO DE DATOS MASIVOS EN BIOMEDICINA COMPUTACIONAL | Víctor Castañeda

# TRANSANTIAGO COMO FUENTE DE DATOS

## LOS DATOS PASIVOS PUEDEN AYUDARNOS A HACER UNA MEJOR GESTIÓN DE LA CIUDAD



### MARCELA MUNIZAGA

Profesora Asociada y Subdirectora del Departamento de Ingeniería Civil Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile. Ingeniero Civil (UCH) y Doctora en Ciencias de la Ingeniería (PUC). Directora Proyecto FONDEF Herramientas avanzadas para la ciudad del futuro. Áreas de investigación: modelamiento del comportamiento de usuarios, recolección y procesamiento de datos.

[mamuniza@ing.uchile.cl](mailto:mamuniza@ing.uchile.cl)

La llegada de Transantiago como sistema integrado de transporte público de Santiago, Chile, a partir de febrero de 2007 fue polémica, algo traumática... en fin, podríamos decir mucho, y ya se ha dicho mucho sobre el tema, pero hay algo sobre lo que no se ha hablado tanto, que es el beneficio colateral de la generación constante de enormes cantidades de datos.

En el Departamento de Ingeniería Civil (DIC) de la Universidad de Chile vimos esa oportunidad y en 2008 comenzamos a trabajar con

los datos en un proyecto PBCT (Programa Bicentenario de Ciencia y Tecnología). El primer desafío fue limpiar y procesar datos de posi-

cionamiento de los más de 6.000 buses del sistema, que cuentan con dispositivos GPS que emiten una señal de posición cada 30 segundos, generando del orden de 80 millones de registros a la semana. Estos permiten observar el movimiento de los buses con un nivel de cobertura y precisión que nunca antes había estado disponible, y generar perfiles de velocidad (Figura 1) [Cortés et al., 2011].

Por otra parte están los datos de transacciones bip!, del orden de 35 millones a la semana, que de por sí contienen información valiosa, al mostrar la distribución temporal de la demanda, pero que además al cruzarlos con la base de datos de posicionamiento hace posible asignar posición a los registros de subida (Figura 2). Esto hace posible obtener la distribución espacio-temporal de la demanda. Mediante una metodología desarrollada en el DIC, que se basa en observar la secuencia de transacciones de una misma tarjeta [Munizaga y Palma, 2012], se estima el paradero o estación de bajada como aquel más conveniente para acceder a la posición de la siguiente subida, dentro de un radio de 500m. Esto se logra para sobre el 80% de las transacciones, abriendo la puerta a una variada gama de análisis, que incluyen construir matrices origen-destino de viajes en transporte público, construir perfiles de carga de buses y Metro (Gschwender et al., 2012), construir indicadores de calidad de servicio, etc. Todas éstas son valiosas herramientas para quienes están encargados de realizar la planificación y gestión

de nuestro sistema de transporte público, que si bien se pueden obtener mediante mediciones, su costo es muy elevado, y su nivel de cobertura espacio-temporal muchísimo menor, dado que existen recursos limitados para realizar este tipo de análisis. Por ejemplo, las matrices origen-destino de viajes tradicionalmente se obtienen de encuestas origen-destino que se realizan a una muestra de la población. En Santiago se realiza una cada diez años, la última se realizó en 2012-2013, con una muestra de alrededor de 18.000 hogares y un costo aproximado de 700 millones de pesos, incluyendo la realización de las encuestas y mediciones complementarias. Los resultados aún no están disponibles, pues con posterioridad al proceso de recolección de datos se requiere un detallado postproceso para filtrar errores y realizar la expansión a la población. Con los datos de transacciones bip! se obtiene información detallada del 80% de los viajes en transporte público, que corresponden a más cuatro millones de viajes diarios de más de dos millones de usuarios (tarjetas bip!), logrando una cobertura espacio-temporal que es imposible alcanzar con datos de encuestas. Mediante postprocesamiento de los datos de transacciones se distingue los transbordos de los destinos de viaje donde los usuarios realizan actividades [Devillaine et al., 2012], y para los usuarios frecuentes (aquellos que viajan al menos cuatro veces a la semana) se identifica la zona de residencia, observando la posición de la primera transacción del día en todos

los días en que la tarjeta es observada. Si éstas tienen coincidencia espacial, se estima que esa zona corresponde al lugar de residencia del usuario (tarjeta). Asimismo, hay otro tipo de análisis que sería interesante realizar, como por ejemplo observar los patrones de viaje y completar información faltante.

Una segunda etapa en el desarrollo de este proyecto es la validación, porque si bien se logra realizar estimaciones de paradero de bajada para sobre el 80% de las transacciones, identificar destinos de actividades para esos viajes, e incluso estimar zona de residencia para los usuarios frecuentes del

sistema, se requiere información exógena para comprobar que esas estimaciones sean correctas. Para realizar la validación hasta ahora se ha contado con una pequeña muestra de validación que proviene de la encuesta origen destino de viajes en Metro, realizada en 2010, en que a una muestra de usuarios se les registra el número identificador de la tarjeta bipl, además de aplicárseles la encuesta origen-destino. Los resultados son positivos, pero insuficientes, debido al reducido tamaño muestral (882 encuestas). Sin embargo, próximamente contaremos con los resultados de la EOD 2012-2013, que es una muestra representativa de la población de Santiago, que podrá ser usada para validación, dado que en ésta también se incluyó una pregunta que pide registrar el número identificador de la o las tarjetas que el encuestado utiliza para viajar.

mejorar la regularidad de los intervalos entre pasadas de buses de un mismo servicio? Dado que ya contamos con información de casi ocho años consecutivos, con amplia cobertura espacio-temporal, es posible observar esos cambios y deducir leyes de comportamiento. Ésta es una etapa fascinante en una disciplina que tradicionalmente se ha enfrentado a la escasez de datos. Hay factores relevantes para las decisiones de los usuarios como por ejemplo la variabilidad del tiempo de viaje o el tiempo de espera, que difícilmente han sido analizados con propiedad debido a la escasez de datos. Ahora cambiamos de paradigma, pasamos de la escasez de datos a la abundancia abrumadora de ellos, y el desafío es utilizarlos adecuadamente en beneficio de la sociedad. El gran desafío es utilizar los datos que se generan automáticamente mediante la operación del sistema, para contribuir a mejorar la gestión de la ciudad, haciendo que todos sus sistemas funcionen de forma más eficiente, amable y sustentable. La invitación está abierta a quienes quieran realizar investigación en esta área, porque aún hay mucho por hacer, y las más diversas disciplinas pueden contribuir a ello, incluyendo por cierto a la Ciencia de la Computación.

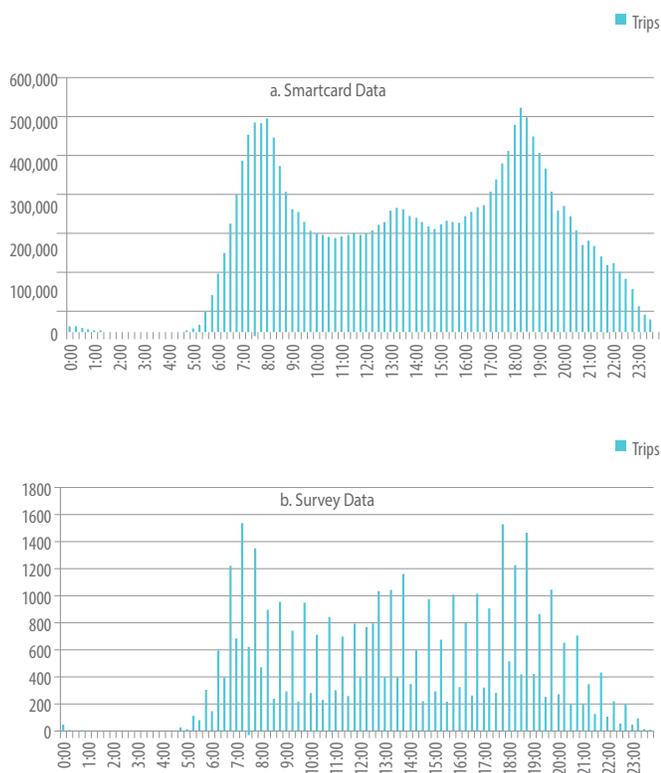
**PROYECTOS QUE HAN CONTRIBUIDO AL FINANCIAMIENTO DE ESTA INVESTIGACIÓN:**

- "Herramientas avanzadas para la ciudad del futuro". Proyecto FONDEF D10I1002 2012-2014.

Otra línea de desarrollo interesante a partir de este proyecto es la modelación. Lo que se obtiene de las transacciones y GPS de los buses es una observación de la realidad actual, algo así como una foto de alta definición, pero para poder aportar al proceso de toma de decisiones, requerimos más que una foto, necesitamos elaborar modelos de comportamiento que nos permitan predecir qué va a suceder con el sistema si aplicamos cambios en él. Por ejemplo, ¿qué va pasar con las velocidades de los buses y los tiempos de viaje de los usuarios si construimos corredores segregados para los buses? ¿Qué efecto tendrá en la demanda



**FIGURA 1.** TRAYECTORIA ESPACIO-TIEMPO DE LOS BUSES DE UN SERVICIO.



**FIGURA 2.** COMPARACIÓN DEL NIVEL DE PRECISIÓN OBTENIDO CON DATOS DE TRANSACCIONES BIP! Y CON DATOS DE ENCUESTA.

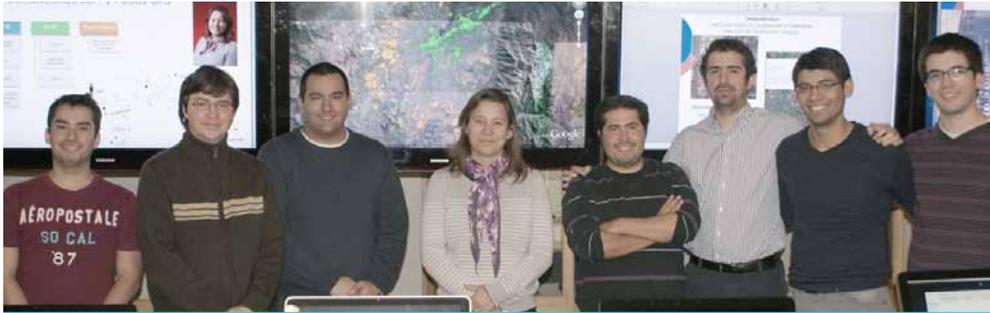


IMAGEN 1.  
DE IZQUIERDA A DERECHA: RICHARD IBARRA\*, ANTONIO GSCHWENDER, CRISTIÁN HERRERA, MARCELA MUNIZAGA, MAURICIO ZÚÑIGA\*, RAMÓN CRUZAT\*, CÉSAR NÚÑEZ, NÉSTOR GALLEGOS (\*GRADUADOS DEL DCC U. DE CHILE).

- "Modeling public transport demand and operation using detailed information from smart-card and AVL data". FONDECYT Regular 1120288. 2012-2014.

- "Transport demand modelling considering quality of service and its impact on demand stability". Proyecto FONDECYT Regular 1090204. 2009-2011.

- "Redes Urbanas" Proyecto Bicentenario de Ciencia y Tecnología PBCTR-19.

- "Instituto Sistemas Complejos de Ingeniería" ISCI (ICM P-05-004-F, CONICYT FBO16).

#### ALUMNOS GRADUADOS EN EL MARCO DE ESTA INVESTIGACIÓN:

- Cristián Herrera. Memoria: Desarrollo de un modelo de elección de ruta en Metro. Ingeniero Civil, mención Transporte. Universidad de Chile (2014).

- Margarita Amaya. Tesis: Análisis de patrones de viaje utilizando datos masivos de transporte pú-

blico. Ingeniero Civil y Magíster en Ciencias de la Ingeniería, mención Transporte. Universidad de Chile (2013).

- Flavio Devillaine. Tesis: Estimación de viajes y actividades en base a sistemas tecnológicos de transporte público. Ingeniero Civil y Magíster en Ciencias de la Ingeniería, mención Transporte. Universidad de Chile (2012).

- Diego Silva. Memoria: Validación exógena de la estimación de paradero de bajada y destino de actividades de usuarios de Transantiago. Ingeniero Civil, mención Transporte. Universidad de Chile (2012).

- Claudio Navarrete. Memoria: Análisis de patrones anómalos en la secuencia de transacciones de pago en el sistema de transporte público de Santiago. Ingeniero Civil, mención Transporte. Universidad de Chile (2012).

- Richard Ibarra. Memoria: Diseño e implementación de un software de cálculo y visualización de perfiles de carga para buses de Tran-

santiago. Ingeniero Civil en Computación. Universidad de Chile (2012).

- Daniel Fischer. Memoria: Obtención de una matriz origen-destino de viajes en transporte público a partir de datos pasivos. Ingeniero Civil Industrial. Universidad de Chile (2010).

- Mauricio Zúñiga. Memoria: Metodología y software para estimar velocidades con datos de Transantiago. Ingeniero Civil en Computación. Universidad de Chile (2010).

- Pamela Mora. Memoria: Generación de datos de patrones de viaje a partir de transacciones BIP. Ingeniero Civil, mención Transporte. Universidad de Chile (2010). ■

## REFERENCIAS

Cortés, C., Gibson, J., Gschwender, A., Munizaga, M.A. y Zúñiga, M. (2011) Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research C* 19(4), 695-707.

Devillaine, F., Munizaga, M.A. y Trepanier, M. (2012) Detection activities of public transport users by analyzing smart card data. *Transportation Research Record* 2276, 48-55.

Gschwender, A., Ibarra, R., Munizaga, M. y Palma, C. (2012) Monitoring Transantiago through enriched load profiles obtained from GPS and smartcard data. *CASPT Santiago, Chile* 23-29 julio.

Munizaga, M.A. y Palma, C. (2012) Estimation of a disaggregate multimodal public transport origin-destination matrix from passive Smart card data from Santiago, Chile. *Transportation Research* 24C(12), 9-18.