



ISIS4822- VISUAL ANALYTICS

Mario Andrés Varón Forero – Juan Sebastián Martínez

200212624-

Noviembre 1 de 2016

Propuesta y Planteamiento del Proyecto

- 1. Nombre de la Propuesta: Visualización para datos probeta en Buses en la ciudad de Bahía Blanca, Argentina.**
2. Caracterización del Problema

La municipalidad de Bahía Blanca, Argentina, utiliza un sistema de AVL (Automatic Vehicle Detection) para la recolección de los datos de posición de los buses durante el recorrido de su ruta. Dichos datos se han recopilado desde el año 2010 hasta la fecha para las distintas rutas que existen en la ciudad. Los dispositivos de medición han sido suministrados e instalados por una empresa privada, no obstante, la municipalidad se ha encargado de la recolección de los datos. Dichos han sido entregados en su totalidad a la Universidad de Nacional del Sur para su análisis y posible visualización. Se ha planteado una alianza entre dicha Universidad y los Andes, con el fin que la visualización se realice dentro del alcance del proyecto de la clase de Visual Analytics del departamento de Ingeniería de Sistemas.

La clasificación de los datos está por secuencia cronológica, incluyendo en orden, el año, el mes, el día y la hora (en formato de 24 horas). Los datos están comprimidos en extensión .kmz, la cual es de tipo de ubicación geográfica para apertura con Google Maps. Su versión descomprimida .kml, contiene una serie de información en lenguaje de etiquetas XML, en donde se incluyen entre otros, la posiciones de latitud y longitud así como el timestamp y otra información de interés.

El dispositivo de AVL, incorpora una unidad de APC (Automatic Passenger Counting System) la cual permite generar un tipo de archivo tipo pass. Este archivo se genera cada vez que un pasajero se sube al bus. Así mismo, la unidad AVL genera cada cierto tiempo un archivo de tipo pcontrol. La frecuencia o la condición por la cual se genera dicho archivo resulta desconocida y se intentará establecer con el estudio de los datos. En este archivo se encuentra entre otros, la placa del vehículo sobre la cual se genera la alerta, la Fecha y la hora, las coordenadas de latitud y longitud, la distancia



que ha recorrido el vehículo durante la jornada de operación, el rumbo, el Estado del vehículo y su velocidad. La estructura de los datos se puede visualizar como se muestra a continuación:

```
e-track-pcontrol-Plaza-Mayo-1433.kml: Bloc de notas
Archivo Edición Formato Ver Ayuda
38.72602333000</coordinates></Point></Placemark><Placemark><styleUrl>#40-NE</styleUrl><ExtendedData><Data
name="Position"><value>-62.22192667000--38.72574167000</value></Data><Data name="Distancia_Hoy"><value>27.64
Kilometros</value></Data><Data name="Distancia_Hoy_Real"><value>0.00 Metros</value></Data><Data name="Lat"><value>-
38.72574167000</value></Data><Data name="Lon"><value>-62.22192667000</value></Data><Data name="Patente"><value>FQD 778
[1433]</value></Data><Data name="Rumbo"><value>Noreste</value></Data><Data name="Estado"><value>Frenado - [Ebus:
Cerrada]</value></Data><Data name="Velocidad"><value>4.0Km/h</value></Data><Data
name="Fecha"><value>25/06/2010</value></Data><Data name="Hora"><value>23:59:22</value></Data><Data
name="Recorrido"><value>Fuera de Recorrido</value></Data></ExtendedData><Point><coordinates>-62.22192667000,-
38.72574167000</coordinates></Point></Placemark><Placemark><styleUrl>#40-NE</styleUrl><ExtendedData><Data
name="Position"><value>-62.22155500000--38.72544000000</value></Data><Data name="Distancia_Hoy"><value>27.69
Kilometros</value></Data><Data name="Distancia_Hoy_Real"><value>0.00 Metros</value></Data><Data name="Lat"><value>-
38.72544000000</value></Data><Data name="Lon"><value>-62.22155500000</value></Data><Data name="Patente"><value>FQD 778
[1433]</value></Data><Data name="Rumbo"><value>Noreste</value></Data><Data name="Estado"><value>Normal - [Ebus:
Cerrada]</value></Data><Data name="Velocidad"><value>22.0Km/h</value></Data><Data
name="Fecha"><value>25/06/2010</value></Data><Data name="Hora"><value>23:59:37</value></Data><Data
name="Recorrido"><value>Fuera de Recorrido</value></Data></ExtendedData><Point><coordinates>-62.22155500000,-
38.72544000000</coordinates></Point></Placemark><Placemark><name>23:59</name><styleUrl>#40-N</styleUrl><ExtendedData><Data
name="Position"><value>-62.22113667000--38.72515333000</value></Data><Data name="Distancia_Hoy"><value>27.69
Kilometros</value></Data><Data name="Distancia_Hoy_Real"><value>0.00 Metros</value></Data><Data name="Lat"><value>-
38.72515333000</value></Data><Data name="Lon"><value>-62.22113667000</value></Data><Data name="Patente"><value>FQD 778
[1433]</value></Data><Data name="Rumbo"><value>Norte</value></Data><Data name="Estado"><value>Frenado - [Ebus:
Cerrada]</value></Data><Data name="Velocidad"><value>0.0Km/h</value></Data><Data
name="Fecha"><value>25/06/2010</value></Data><Data name="Hora"><value>23:59:52</value></Data><Data
name="Recorrido"><value>Fuera de Recorrido</value></Data></ExtendedData><Point><coordinates>-62.22113667000,-
38.72515333000</coordinates></Point></Placemark></Folder></Folder></Document></kml>
```

Existe un archivo tipo REC que contiene la posición de latitud y longitud para toda la ruta recorrida por el bus y finalmente un archivo de tipo STOP que se genera en el momento de detención de un bus.

En la medida en que se puede obtener de los datos, tanto la ruta, como la cantidad de pasajeros, así como la velocidad media del bus y el tipo de alarmas que se generan, el alcance del proyecto debe estar involucrado directamente con estos datos o con la derivación que se pueda hacer de ellos.

3. Objetivos:

- Visualizar la información descriptiva de tráfico obtenida de los datos probeta con el fin de obtener medidas de desempeño (i.e tiempo de viaje o costo de viaje) que permitan construir un sistema ATIS o realizar cambios en la planeación de las rutas de transporte.
- Visualizar la información del APC (Automatic Passenger Counting System) con el fin de optimizar el costo de operación de los buses



públicos en la ciudad de Bahía Blanca, bien, modificando su horario de funcionamiento o el trazado de su ruta .

- Preprocesar y Tratar los datos para obtener medidas estadísticas poblacionales que permitan caracterizar el comportamiento de los buses y en menor medida, del tráfico en la ciudad de Bahía Blanca.

4. Caracterización de la Visualización

What:

Datos geográficos (geometry-spatial) con varios atributos cuantitativos secuenciales, entre esos velocidad de desplazamiento, cantidad de pasajeros, distancia recorrida, rumbos y estado del vehículo clasificados por datos categóricos, placa del bus.

Why:

Descubrir la distribución de los datos de velocidad de desplazamiento promedio por día por ruta de transporte público.

Explorar los Outliers de velocidad para identificar eventos de congestión no recurrente o recurrente.

Identificar los puntos geográficos de las rutas de los buses donde su velocidad de desplazamiento es menor.

Descubrir la cantidad de pasajeros que viajan por ruta por día.

Identificar las rutas con menor/mayor cantidad de demanda por trayecto origen-destino.

How:

Para la Tarea 1, 2, 3, 4 y 5: Map con formas (shapes) que usan la geometría de los datos del AVL organizados en Horizon Charts para cada tipo de ruta.

5. Estado del Arte

La definición más adecuada para los sistemas de información al viajero avanzadas (ATIS) sería "la aplicación sistemática de las tecnologías de información y comunicación para la recopilación de datos relacionados con los viajes y el procesamiento y la entrega de información al viajero" (McQueen, 2002). Este concepto implica claramente tanto a los viajeros en las redes de transporte público como aquellos que hacen uso del privado.



Zhang (Zhang, y otros, 2010), y Adler (Adler & Blue, 1998) clasifican los sistemas de información al viajero (TIS) en dos generaciones: la primera consiste en proveer la guía de ruta al viajero por medio de señales de mensajes variables (VMS), mientras que la segunda generación (ATIS) se puede definir como un proceso dinámico para la guía de ruta mediante el uso sistemático nuevas tecnologías que recojan la información del tráfico en tiempo real. Así mismo, Koppelman (Koppelman & Pas, 1980) y Kanninen (Kanninen, 1996) han argumentado que proveer a los viajeros con información relevante acerca de las opciones sobre su viaje tiene el potencial de cambiar sus comportamientos en formas que son beneficiosas para la mejorar la eficiencia en el uso de la red de transporte. Sin embargo, esta perspectiva es bastante limitada como fundamento principal del uso y despliegue de sistemas ATIS. Si se incorpora en el análisis la teoría microeconómica del consumidor (Samuelson, 1947), serán las evaluaciones exhaustivas de las alternativas (opciones de ruta), así como la exploración de atributos para cada una y el cálculo de la utilidad del intercambio entre ellas lo que tiene valor para el viajero. Por tanto y de aquí en adelante debe entenderse por un sistema ATIS como : "la aplicación sistemática de las tecnologías de información y comunicación con el fin de proporcionar la mejor guía de ruta a través de la recopilación de datos relacionados con los viajes en tiempo real, así como su procesamiento y la entrega de información al viajero para que puedan evaluar de manera exhaustiva cada uno de los atributos de las alternativas provistas y en consecuencia realizar los respectivos intercambios entre ellas para maximizar su utilidad (por ejemplo, reducir el tiempo de viaje) ".

Con el fin de seguir la tesis delineada, hay varios atributos que pueden ser proporcionados por un sistema ATIS para los viajeros públicos en las redes de transporte. Estos atributos se pueden dividir en dos clases diferentes, el primer tipo de atributos de viajes están relacionados con la congestión no recurrente y los segundos, a la congestión recurrente. McGroarty (McGroarty, 2010) ha definido la primera como "demoras de tráfico inesperadas causadas principalmente por accidentes e incidentes, averías de vehículos, actividades de construcción de carreteras, eventos especiales, eventos climáticos extremos, etc" por lo que los atributos relacionados específicamente con este tipo de eventos y que se proporcionarían para el viajero, serán el tipo, la localización, la duración y su reducción de la capacidad en el arco donde ocurrió el incidente, así como su impacto en el tiempo de viaje esperado . El viajero podrá estar interesado en la recoger esta información antes del viaje y durante el momento del viaje. Cabe mencionar que las únicas decisiones



posibles ante esta información por parte del viajero, es el cambio de la ruta en el transporte público, la modificación de la hora de salida o eventualmente el cambio de modo de transporte (i.e cambiar el bus por el metro).

Por otra parte, la congestión recurrente se asocia con el diagrama fundamental en la modelación macroscópica, que se representa por el flujo, la densidad y la velocidad (Roess, Prassas, & McShane, 2004). Los atributos proporcionados a los viajeros, en este tipo de circunstancias, se basan en cálculos directos o indirectos de estas variables, y se expresan por medio del tiempo de viaje, velocidad media, volumen y costo de viaje. Para que el viajero se familiarice con el comportamiento del tráfico y pueda construir una percepción informada, la recolección de la información debe hacerse en el pre-viaje, durante el momento del viaje y en después de terminado viaje. (Chorus, Molin, & Van Wee, 2005). De la misma forma que para la congestión no recurrente, las opciones quedan delimitadas al cambio de ruta, la modificación en la hora de salida o el cambio en el modo de transporte. No obstante, mientras que para la primera opción dicha decisión afecta el viaje que se hará a continuación, en este segundo caso, la decisión modifica un hábito de viaje y se perpetúa a lo largo del tiempo, hasta que llegue nueva información que permita evaluar nuevamente, la función de utilidad del viajero.

Sin embargo, hay información de tráfico que no se derivan ni la congestión recurrente, ni tampoco de la no recurrente, Torres (Torres, 2008), McQueen (McQueen, 2002) y Veneziano (Veneziano, 2010), los han descrito en detalle, e incluyen entre otros, estado meteorológico, puntos de interés, puntos de fiscalización del tráfico vehicular, imágenes de CCTV y otros que son menos habituales. De hecho, existe evidencia que algunos de estos parámetros modifican drásticamente variables como la velocidad o el flujo vehicular, no obstante de una u otra forma se reflejan en las características de los atributos referidos a la congestión recurrente y no recurrente. En este sentido, Wunderlich (Wunderlich, 1996) ha calculado que las condiciones meteorológicas adversas pueden reducir hasta un 75% la capacidad de la red. Sin embargo, ninguno tipo de esta información adicional puede considerarse como atributos de viaje, ya que no afecta a la utilidad individual personal y por lo tanto no puede ser objeto de intercambio entre las distintas alternativas provistas por el sistema ATIS. (Innocenti, Lattarulo, & Pazienza, 2009)

El conjunto de opciones de viaje relacionados con los atributos mencionados anteriormente, incluye, cancelar el viaje, modificar el modo, cambiar la hora de salida y finalmente evaluar las alternativas de elección de ruta y escoger



una. (Khattak, Polydoropoulou, & Ben-Akiva, 1995). Cuando el viajero adopta cualquiera de estas opciones, está realizando directamente una evaluación entre los distintos atributos, esperando aumentar su utilidad de su decisión escogida. Teniendo en cuenta este hecho y la definición propuesta para un ATIS en este documento, un ATIS comercial es el único medio que pudiera ofrecer a los viajeros la información para que pueda construir y medir la utilidad objetiva y tomar por lo menos una de las opciones previstas anteriormente. Para todos los efectos de este documento, se entiende que la maximización de la utilidad está en función de reducir su tiempo de conmutación entre su punto de origen y el destino de su viaje.

En la medida en que se obtienen los datos de velocidad y tiempos de viaje para cada una de las rutas, la información provista puede soportar un ATIS de tipo descriptivo para los viajeros, así como permitir la intervención por medio de política públicas que permitan optimizar dichos tiempos, con su consecuente beneficio en la productividad de la ciudad.

Finalmente los datos de pasajeros por ruta se utilizan principalmente para verificar la consistencia en la tarifa del transporte público, siempre y cuando, dicha sea fijada por la municipalidad. Así mismo, tales datos son útiles para determinar en el nivel macroscópico la matriz origen-destino, así como la determinación de la factibilidad de ciertas rutas de transporte público, bien cuándo su cantidad de pasajeros esté por debajo de su nivel de factibilidad o en su defecto cuando se encuentren por encima, obligando a la entidad a adoptar más rutas para un mismo viaje con idénticas coordenadas de origen-destino.

6. Método propuesto

De acuerdo a las tareas identificadas para la solución del problema, se propuso una visualización compuesta por dos conjuntos de “Horizon Charts”, el primero permitirá visualizar el comportamiento de las velocidades por cada una de las rutas de buses en Bahía blanca en función del tiempo para un cierto día, mes y año. Otro gráfico de la misma naturaleza estará destinado a presentar la cantidad de pasajeros por cada bus a lo largo del tiempo; nuevamente para un cierto día, mes y año.

Estas visualizaciones se proponen ya que se ha de tratar datos temporales, que presentan cantidades numéricas, es por esto que los “Horizon Charts” presentan los datos sobre una misma escala lineal para un cierto día y utilizan el color como un canal efectivo para presentar rangos en los cuales los datos están presentes. Estos son: velocidades superiores e inferiores a la velocidad promedio de dicho día, y cantidad de personas superiores e inferiores al



promedio de personas que transporta el bus en el día. Es por esto que la visualización tiene una alta efectividad a la hora de soportar las tareas de descubrir las distribuciones y “outliers” de velocidades y cantidad de personas durante los días.

Más aún, se propone como interacción la posibilidad de seleccionar puntos específicos de cada uno de los “Horizon Charts” para visualizar, en un mapa de la ciudad, la ruta seleccionada y el punto geográfico correspondiente que corresponde a la velocidad o cantidad de personas que generó interés en el usuario.

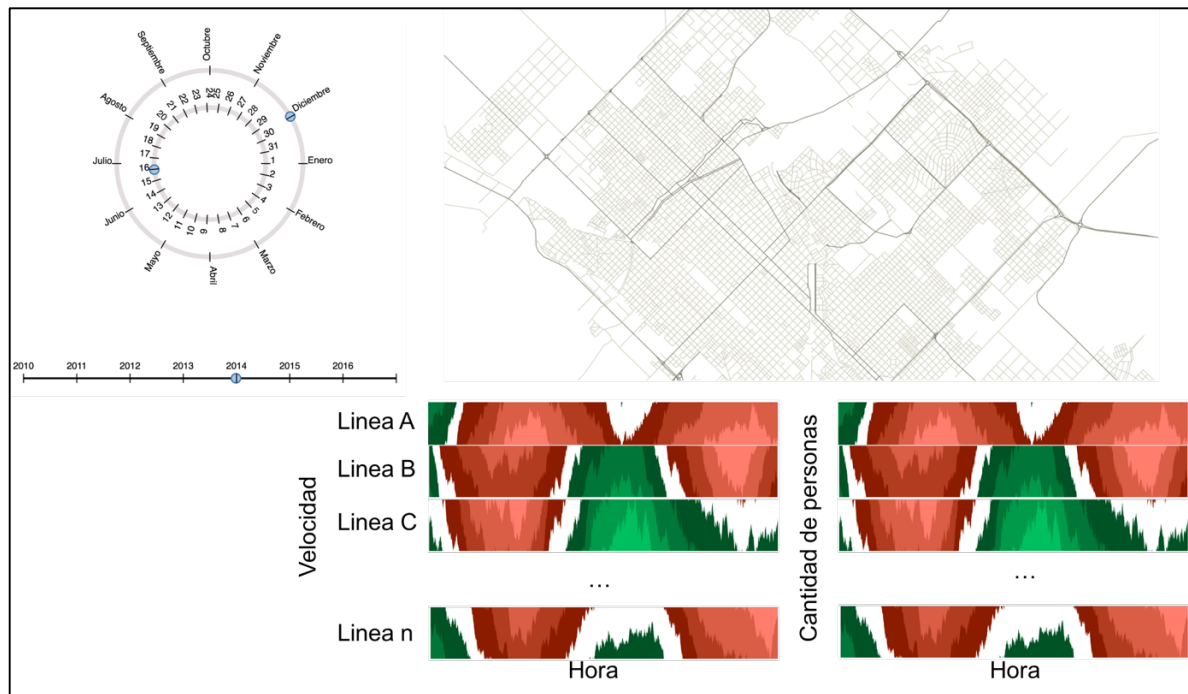
Por último, dado que las visualizaciones tienen una granularidad horaria por día, los datos son filtrados por año, mes y día a partir de dos controles circulares y un control lineal. Los primeros dos permiten filtrar el mes y el día de acuerdo a su naturaleza cíclica; el último permite seleccionar el año y es un control lineal debido a que es un dato lineal.

7. Primeros resultados

De acuerdo a las conversaciones con el cliente, ha sido posible obtener una totalidad de 150GB de datos de buses durante 6 años que fueron suministrados para el proyecto, los datos no han sido visualizados antes por lo cual actualmente no se cuenta con herramientas de visualización para su análisis.

Un análisis preliminar sobre los datos reveló que es necesario realizar un proceso de extracción, transformación y carga de los datos (ETL por sus siglas en inglés). Como se ha observado, una gran cantidad de archivos presentes en la información suministrada no contienen datos útiles, por lo cual se espera que la densidad de los datos a utilizar en la visualización disminuya considerablemente.

En cuanto a las visualizaciones, la siguiente figura muestra un mockup del diseño presentado al cliente, tal y como se describió en la sección anterior.



8. Retroalimentación por parte del Experto

A partir de la propuesta presentada al experto, se obtuvo una retroalimentación positiva por parte del cliente, así como una serie de sugerencias con respecto a la visualización y con respecto a trabajos futuros.

En primer lugar, el cliente expresó su satisfacción con la visualización y el hecho de que cumple efectivamente con las tareas planteadas desde la primera reunión. Afirmó además que, de generar “insights” completamente nuevos para el usuario, las vistas pueden llegar a evolucionar para cumplir nuevas tareas. Como recomendaciones se determinó que es mejor visualizar los datos en intervalos de media hora, en lugar de horas y de tratar de visualizar las rutas de los buses en sus respectivas polilíneas, en lugar de los puntos geográficos.

Por otro lado, dado que una de las tareas corresponde a descubrir las características de los pasajeros, se propone estimar una matriz "origen-destino" para las bajadas de los pasajeros, con el fin de obtener información sobre el comportamiento de las personas al usar el transporte público. También, el cliente sugiere ubicar geográficamente las paradas de los buses en el mapa.



9. Bibliografía

- Adler, J. L., & Blue, V. (1998). Toward the design of intelligent traveler information systems. *Transportation Research Part C: Emerging Technologies*, Vol. 6, No. 3, pp. 157 – 172.
- Chorus, G., Molin, E., & Van Wee, B. (2005). Use and Effects of Advanced Traveller Information Services (ATIS): A Review of the Literature. *Transport Reviews*, Vol. 26, No. 2, 127–149. .
- Innocenti, A., Lattarulo, P., & Pazienza, M. (2009). An Experimental Analysis of Travel Mode Choice. *Società Italiana di Economia dei Trasporti e della Logistica - XI Riunione Scientifica* –. Trieste.
- Kanninen, J. B. (1996). Intelligent transportation systems: an economic and environmental policy assessment. *Transportation Research Part A*, 30, pp. 1-10.
- Khattak, A., Polydoropoulou, A., & Ben-Akiva, M. (1995). Modeling Revealed and Stated Pretrip Travel Response to ATIS. *Transportation Research Part*, 1537, pp. 46-54.
- Koppelman, F., & Pas, E. (1980). Travel choice behaviour: models of perceptions, feelings, preference, and choice,. *Transportation Research Record*, 765, pp. 26-33.
- McGroarty, J. (2010). Recurring and Non Recurring Congestion: Causes, Impacts and Solution. *Neihoff Urban Studio*, pp 2-6.
- McQueen, B. (2002). *Advanced Traveler Information Systems Intelligent Transportation systems*. Norwood: Artech house.
- Roess, R., Prassas, E., & McShane, W. (2004). *Traffic Engineering*. Washington: Pearson.
- Samuelson, P. (1947). *Foundations of economic analysis*. Harvard University Press: Cambridge.
- Torres, N. (2008). *ATIS Final report*. Hartford: Department of Civil and Enviromental Engineering University of Hartford.
- Veneziano, D. (2010). *Rural Traveler Information Needs Assessment and Pilot Study*. Sacramento: California DOT.
- Wunderlich, K. (1996). *An Assessment of Pre-Trip and en route ATIS Benefits in a Simulated Regional Urban Network*. Orlando: ITS America.
- Zhang, L., Li, J., Zhou, K., Gupta, S., Li, M., Zhang, W.-P., . . . Misener, J. (2010). Design and Implementation of a Traveler Information Tool with Integrated Real-Time transit Information and Multi Modal Trip Planning. *Transport Research Board 91th Meeting*. Washington.