

METODOLOGÍA PARA LA FORMULACIÓN DE INDICADORES DE CALIDAD DE SERVICIO EN EL TRANSPORTE PÚBLICO A PARTIR DEL USO DE DISPOSITIVOS INERCIALES

Hernán Catalán S., Pontificia Universidad Católica de Chile (hfcatala@uc.cl)

Juan Carlos Herrera M., Pontificia Universidad Católica de Chile (jch@ing.puc.cl)

RESUMEN

Actualmente el sistema de transporte público de Santiago emplea un Indicador de Calidad de Atención al Usuario (ICA) que aborda superficialmente la calidad de servicio y de una manera muy subjetiva. El objetivo de esta investigación es formular una nueva metodología que utilice la información de acelerómetros y giroscopios para clasificar la forma de conducción del chofer del bus, mediante una máquina de soporte vectorial (SVM). Los resultados preliminares muestran que es factible usar SVMs con este propósito, pero que existen fallas que pueden asociarse a la metodología de pasajero incógnito utilizada. Otra fuente de error puede ser el uso de una muestra desbalanceada.

Palabras claves: *dispositivos inerciales, transporte público, forma de conducción.*

ABSTRACT

Currently, Santiago's public transit system employs a User Service Quality Indicator (ICA) to measure the quality of service offered in a superficial and very subjective manner. The main objective of this research is to formulate a new methodology that uses the information provided by accelerometers and gyroscopes to classify drivers' behavior. This is achieved by means of a Support Vector Machine (SVM). Preliminary results show that it is feasible to use SVM for this purpose, but there are flaws that can be associated with the use of an incognito passenger. Another source of error may be the use of an imbalanced sample.

Keyword: *inertial measurement unit, public transport, driving style.*

1. INTRODUCCIÓN

El sistema de transporte público Transantiago ha sido fuertemente criticado desde que se puso en marcha hace 10 años. Según un estudio encargado por la Directorio de Transporte Público Metropolitano (DTPM, 2016), el 2016 los usuarios calificaban a Transantiago con nota 4,3 (escala de 1 a 7, con 4 siendo nota suficiente). Además, del total de reclamos del año 2016, el 27% corresponden al manejo y comportamiento del conductor, siendo la segunda razón más frecuente de reclamos (después de la no detención en paraderos).

El sistema ha sufrido múltiples modificaciones, algunas más radicales que otras, con el objetivo común de ofrecer un servicio de calidad a sus usuarios. Si bien los contratos consideran desde los primeros años indicadores operacionales con altos porcentajes de cumplimiento (ICF e ICR), recién desde 2012 entra en funcionamiento el Índice de Calidad de Atención al Usuario (ICA). Este considera 14 atributos asociados al servicio entregado al usuario tanto como el que va a bordo del bus como el que espera en el paradero.

La metodología de medición del ICA es de pasajero incógnito. Es decir, una persona que se comporta como cualquier pasajero y que evalúa cada atributo de manera binaria: cumple (1) o no cumple (0). Además, los contratos estipulan que cada mes se medirá el indicador aleatoriamente en el 25% de los buses de una empresa y durante 20 minutos por cada bus evaluado (DTPM, 2013).

Claramente la forma en que se mide este indicador acarrea subjetividad y, principalmente, no existe continuidad en la toma de datos, por lo que es difícil para los operadores detectar los focos de malos resultados. De hecho, durante el 2016 el promedio de cumplimiento fue de 78% con una exigencia mínima de 85%. Cuando no se cumple este umbral (DTPM, 2013), los operadores deben pagar multas que van desde las 200 UF (aproximadamente US\$ 8.000).

Entre los atributos a evaluar, el atributo a02: “El conductor conduce sin frenazos ni movimientos bruscos” llega a un 59% de cumplimiento. Si bien la empresa operadora con mejores resultados en este atributo llega a un cumplimiento del 73%, la empresa con menor cumplimiento llega al 53%.

Es claro que la forma de conducción de los choferes del Transantiago no es adecuada, pero bajo la metodología actual no se incentiva a los operadores a mejorarla. Como la muestra evaluada es tan pequeña temporalmente, la medición puede no ser representativa del real cumplimiento del operador. Además, los resultados no se entregan por bus medido, es decir, no se puede identificar a los choferes que necesitan capacitación en la conducción. Por otro lado, no existe un incentivo económico fuerte que ayude a mejorar la situación.

En la última década y con el desarrollo de nuevas tecnologías, el uso de dispositivos iniciales (IMU por su sigla en inglés) se ha masificado debido a su utilidad y bajo costo. Un dispositivo inercial es aquel que posee: giroscopio, acelerómetro y GPS. Permite capturar cuantitativamente el movimiento, aceleración y vibraciones experimentadas por un objeto.

El objetivo general de este trabajo es utilizar los datos recolectados por dispositivos iniciales para formular una nueva metodología para el cálculo del atributo de conducción del ICA de

Transantiago. Esto se logra al clasificar la forma de conducción tomando en cuenta las vibraciones, aceleraciones y movimientos que experimentan los pasajeros a bordo de los buses. El beneficio de esta nueva metodología es que permite calcular los atributos de la conducción del ICA de manera simple y se aplica de igual manera a todos los buses del sistema.

Este trabajo se estructura de la siguiente manera: el capítulo 2 consiste de una revisión bibliográfica que muestra distintas aproximaciones para la cuantificación de la calidad de servicio y la comodidad de un transporte público. El capítulo 3 es un marco teórico de la investigación que busca explicar la teoría detrás de una máquina de vectores de soporte. El capítulo 4 explica la metodología empleada para recabar datos con dispositivos iniciales con los cuales se calibrará la máquina de aprendizaje. El capítulo 4 aborda la modelación y los rendimientos que tienen los modelos al predecir los sets de testeo generados aleatoriamente. Finalmente, el capítulo 5 habla de las conclusiones de la investigación y los pasos a seguir.

2. REVISIÓN BIBLIOGRÁFICA

La calidad del servicio impacta positivamente en la demanda del transporte público (Paultey *et al.*, 2006), por lo que una mejora en ella podría ayudar a disminuir los viajes en auto que se realizan en Santiago. Según el denominado “círculo vicioso del transporte público”, la congestión vehicular es uno de los factores que reduce el uso de los buses favoreciendo otros medios de transporte como el automóvil (Ortúzar & Willumsen, 2011).

Cuando se quiere medir la calidad de un servicio de transporte público, se suele buscar los atributos que la explican. Los modelos desarrollados por Parasuraman *et al.*, 1988; Hensher, 2003 han sido un gran avance en la materia, pero sus atributos son más abstractos que los usados en modelos más recientes. Esto lleva a que sea difícil de cuantificar cada atributo, por ende, replicar dicho modelo en otras circunstancias es muy costoso.

Otros estudios intentan descifrar la importancia que el usuario le entrega a cada atributo. En Grujić *et al.* (2014), se evalúan 24 atributos y se genera un mapa de cuadrantes según nivel de satisfacción e importancia. Dentro de las conclusiones se puede apreciar que una conducción segura es el atributo más importante, pero para el caso estudiado tenía un buen nivel de satisfacción. En de Oña *et al.* (2015), la comodidad se mide como la disponibilidad de asientos o de espacio para ir de pie. Además, existe un atributo asociado a una conducción apropiada. En este modelo, la conducción no fue significativa en comparación a la limpieza del bus o la iluminación en paraderos. Similar resultado fue encontrado por dell’Olio *et al.* (2011).

En Sánchez-Flores y Romero-Torres (2010), realizan un estudio en un corredor de transporte público de la ciudad de Toluca, México, que es servido por más de un recorrido de bus. Concluyen que la forma de conducción es el atributo más significativo para los usuarios al momento de escoger un recorrido. Queda claro que no hay consenso sobre cuáles atributos impactan más en la percepción de la calidad de servicio, si no que es un problema particular a cada sistema y ciudad.

En Nuñez *et al.* (2013) se considera la comodidad como un atributo relevante para describir la calidad de servicio de un sistema de transporte público, pero se descarta del modelo final debido a la dificultad de cuantificar la comodidad con datos pasivos. A una similar conclusión se llega en

Rizzi *et al.* (2005). Esto demuestra la necesidad de buscar alternativas que permitan superar esta barrera.

En Han (2013), se utiliza el IMU a bordo de los trenes que recorren la línea Gyeongbu entre Seoul y Busan, Corea del Sur, para predecir el instante en que el vagón se inclina (por ej. debido a una curva). El objetivo de estos modelos predictivos es contrarrestar la inclinación antes de que esta ocurra y así evitar que los pasajeros experimenten aceleraciones fuera de la norma. Estos modelos predictivos resultan mejor que acciones reactivas.

En Zhang *et al.* (2014), utilizan dispositivos capaces de capturar el ruido, la temperatura, la vibración y la aceleración experimentada a bordo de 13 líneas de bus del sistema de transporte público de la Nanjing, China. Dentro de su trabajo, definen $0,6 \text{ m/s}^3$ como el umbral del *jerk* (tasa a la cual cambia la aceleración), lo que se traduce en que valores mayores implican cambios en la aceleración incómodos para el usuario. Además, los autores se basan en la norma ISO 2631 para clasificar el nivel de comodidad experimentado según la vibración medida en RMS (*Root Mean Square*, m/s^2), como se aprecia en la Tabla 1. En Figueiredo y Silva (2015); Castellanos y Fruett (2013) se utiliza la norma ISO con el mismo propósito.

Tabla 1. Definición Índice de Comodidad. Fuente: Castellanos y Fruett (2013).

Aceleración (G)	Percepción
$< 0,032$	Cómodo
$0,032 - 0,057$	Un poco incómodo
$0,057 - 0,092$	Suficientemente incómodo
$0,092 - 0,145$	Incómodo
$0,145 - 0,229$	Muy incómodo
$> 0,229$	Extremadamente incómodo

Finalmente, en Bodini *et al.* (2013) se adapta al transporte público terrestre, la norma ENV 12299 que define indicadores de comodidad para el sistema ferroviario europeo. Específicamente, se define el indicador para personas de pie (N_{VD}) de la siguiente manera:

$$N_{VD} = 3 \cdot \sqrt{16 \cdot a_x^2 + 4 \cdot a_y^2 + a_z^2 + 5 \cdot a_y} \quad (1)$$

donde a_x es la aceleración longitudinal, a_y es la aceleración lateral y a_z es la aceleración vertical. A mayor valor de N_{VD} , mayor es la incomodidad percibida por los usuarios. El N_{VD} es calculado y almacenado en un servidor cada segundo con una posición asociada (gracias al GPS presente en el IMU), para luego analizar espacial y estadísticamente la información recolectada.

Se puede apreciar como en la literatura no hay una manera específica de medir la comodidad, se refuerza la importancia de la conducción en la percepción de calidad por parte de los usuarios. Además, medir las vibraciones en los distintos ejes puede ser una metodología apropiada para cuantificar la conducción en el transporte público.

3. MÉTODOS DE CLASIFICACIÓN: SVM

Debido a que se desea clasificar la forma de conducción de manera binaria (buena o mala) dada la información recolectada por dispositivos iniciales, proponemos el uso de métodos de clasificación. Estos métodos identifican a cuál categoría pertenece una nueva observación, en base a un set de datos de entrenamiento de los cuales sí se conoce su categoría. Uno de los métodos de clasificación más simples y con buen rendimiento es *Support Vector Machines* (Hsu *et al.*, 2016; Ebrahimi *et al.*, 2017). Por lo tanto, es el método escogido para el desarrollo de este trabajo.

Support Vector Machines es una máquina de aprendizaje que se basa en la teoría de aprendizaje estadístico y en el principio de minimización de riesgo estructural. Es útil para regresiones (análisis de series de tiempo) y reconocimiento de patrones (problema de clasificación), entre otros problemas.

SVM tiene un buen rendimiento al resolver problemas de clasificaciones con muestras desbalanceadas, no lineales y de dimensión superior. Además, tiene la ventaja de evitar problemas de sobre-entrenamiento que reducen la generalización del modelo, es decir, aprende solo el set de entrenamiento (Wang *et al.*, 2015).

En este trabajo se tiene un problema de clasificación, por lo que en esta sección se explicará la formulación teórica básica que utiliza SVM para resolver este problema (Boser *et al.*, 1993; Cortes y Vapnik, 1995).

Un problema de clasificación requiere separar los datos recolectados en un set de entrenamiento y uno de testeо. El primero es aquel que utilizará la máquina de aprendizaje para calibrar el modelo y el segundo, para medir el rendimiento de la predicción. Cada instancia del test de entrenamiento contiene atributos (los datos entregados por el IMU) y una etiqueta de clasificación (conducción buena o mala). El objetivo de SVM es entregar un modelo calibrado capaz de predecir correctamente la etiqueta de clasificación del set de testeо considerando solamente los atributos.

Dado un set de entrenamiento (x_i, y_i) , con $i = 1 \dots L$; donde $x_i \in \mathbb{R}^n$ e $y_i \in \{1, -1\}^L$, SVM debe resolver el siguiente problema de optimización:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^L \xi_i \\ \text{s. a.} \quad & y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned} \tag{2}$$

Los vectores de entrenamiento x_i son llevados por la función ϕ a un espacio de dimensión superior. El SVM encuentra un hiperplano lineal que sea capaz de separar estos vectores en el espacio de dimensión superior con el máximo margen posible, tomando en cuenta la etiqueta de clasificación buscada. El parámetro positivo C es la penalización aplicada al error de clasificación. Además,

$$K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j) \quad (3)$$

es la denominada función kernel. Las más utilizadas en la literatura son:

- Lineal: $K(x_i, x_j) \equiv x_i^T x_j \quad (4)$
- Radial Basis Function (RBF): $K(x_i, x_j) \equiv \exp(-\gamma \|x_i - x_j\|^2) \quad (5)$

con $\gamma > 0$, parámetro de la función kernel RBF.

3.1 Caso de la muestra no balanceada

Un muestra es no balanceada cuando la cantidad de datos positivos (aquellos que el modelador desea identificar, es decir, mala conducción en nuestro caso) es mucho menor a la cantidad de datos negativos (buena conducción). Debido a que las maniobras de mala conducción son instantáneas, la cantidad de alertas producidas sea muy baja, es decir, nuestra muestra es no balanceada. Esto es un problema, ya que en casos extremos (datos positivos menores al 5% de la muestra) el modelo clásico de SVM no aprende la clase positiva, ya que equivocarse está dentro del margen de error aceptable para cualquier método de clasificación. A pesar de ello, SVM tiene una ventaja ante este problema.

En la ecuación (2) se puede ver el problema de optimización que resuelve SVM. La constante C es la penalización por equivocarse al clasificar, por lo que para abordar el problema de muestras no balanceadas, esta constante se divide en el número de clases de la muestra. En otras palabras, se tiene un parámetro C_i para cada clase. Al asignar parámetros tal que $C_1 > C_{-1}$, se fuerza al SVM a aprender la clase positiva.

3.2 Elección de Parámetros

Una de las ventajas de usar SVM es que el modelo requiere de pocos parámetros para ser calibrado. El número exacto depende de la función kernel escogida. Para el caso de kernel lineal (ecuación 4), se requiere calibrar un parámetro: C . Para el kernel RBF (ecuación 5), C y γ .

Particularmente, como la muestra es no balanceada, el parámetro C depende del número de clases. Es decir, para el kernel lineal se requiere calibrar dos parámetros C_1 y C_{-1} ; para el kernel RBF, C_1 , C_{-1} y γ .

Cuando la cantidad de parámetros es superior a uno, se deberán probar múltiples combinaciones de parámetros hasta encontrar la mejor. Esto se determina luego de realizar una validación cruzada dentro del set de entrenamiento que entrega la exactitud (*exactitud*) del modelo. El tipo de validación cruzada es *k-fold* y consiste en separar la muestra en k grupos. Secuencialmente, se deja un grupo fuera como set de testeo y se entrena con los $k - 1$ restantes.

Para facilitar la búsqueda de la mejor combinación, se utiliza el algoritmo de *grid-search*. Es ampliamente aceptado usar incrementos exponenciales hasta encontrar la vecindad con mejores resultados.

3.3 Medidas de Rendimiento para SVM

El rendimiento de un modelo SVM se mide según distintos indicadores evaluados sobre el set de testeo. Los utilizados en este trabajo se basan en la matriz de confusión (ver Tabla 2) y son los siguientes:

$$Exactitud = \frac{VP + VN}{población\ total} \quad (7)$$

$$Error\ Tipo\ I = \frac{FN}{VP + FN} \quad (9)$$

$$Exhaustividad = \frac{VP}{VP + FN} \quad (11)$$

$$Error\ Tipo\ II = \frac{FP}{VN + FP} \quad (8)$$

$$Precisión = \frac{VP}{VP + FP} \quad (10)$$

$$F1\ Score = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \quad (12)$$

Tabla 2. Matriz de Confusión.

Población Total	Predicción Positiva	Predicción Negativa
Condición Positiva	Verdadero Positivo (VP)	Falso Negativo (FN)
Condición Negativa	Falso Positivo (FP)	Verdadero Negativo (VN)

4. METODOLOGÍA

4.1 Herramientas Utilizadas

Una de las ventajas que tienen los dispositivos iniciales (acelerómetros, giroscopios, GPS) es que se encuentran disponibles a todas las personas, ya que están presentes en cada *smartphone*. Estos equipos usan IMUs para controlar la posición de la pantalla (vertical, horizontal), como una extensión de realidad aumentada en juegos (manubrio en un juego de carreras), etc. Por lo general, la información de los IMUs está disponible para desarrolladores de aplicaciones de *smartphone*.

Para este trabajo se utilizó un iPhone 6S de 32 GB, modelo A1633. Este posee un acelerómetro triaxial Bosch Sensortec 367LA, un giroscopio de 6 ejes InvenSense MP67B y un GPS Qualcomm WTR1605L. Adicionalmente, para recuperar la información de los dispositivos se utilizó la aplicación “SensorLog 1.9.2” (Thomas, 2017) disponible en la AppStore. Dentro de las ventajas que tiene esta aplicación es el uso de una interfaz amigable que muestra en vivo gráficos con los datos de cada chip incorporado en el iPhone. Además, almacena la información en archivo .csv con una frecuencia escogida por el usuario que va desde 1Hz a 100 Hz.

Por otro lado, para registrar las instancias de conducción inapropiada se utilizó un iPad mini de 32 GB, modelo A1454 en conjunto de la aplicación “Counter – Advanced Tally Counter 7.1” (Mikrasya, 2016) disponible en la AppStore. Esta aplicación, como su nombre lo sugiere, es un contador que almacena la información en un archivo .csv con el *timestamp* de cada instancia.

4.2 Recolección de datos

Para la recolección de datos se escogió una ruta que consta de 4 tramos de similar duración y distancia. Una descripción detallada de cada uno se puede encontrar en la Tabla 3. Además, en la Figura 1 se puede apreciar un mapa con cada tramo.

Tabla 3. Detalle de cada tramo utilizado para la recolección de datos.

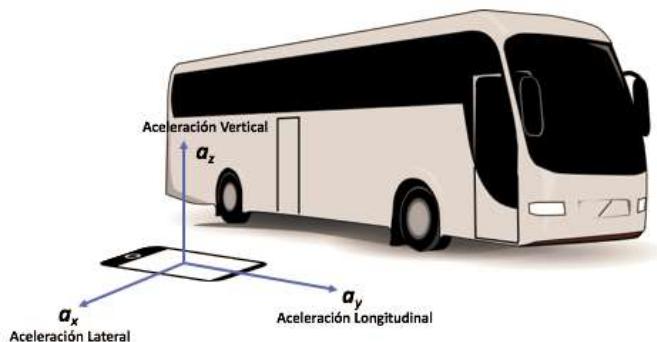
Recorrido	Desde – Hasta	Distancia (km.)	Tiempo (min.)	Nº Datos	Operador	Ránking ICA
213E	Metro Mirador – Plaza Baquedano	8,8	21	63.503	STP Santiago S.A.	1°
210V	Plaza Baquedano – Metro Pedrero	8,4	22	67.319	Subus Chile S.A.	7°
102	Metro Pedrero – Av. La Florida con Rojas Magallanes	5,5	11	32.599	Alsacia S.A.	2°
712	Av. La Florida con Rojas Magallanes – Av. América Vespucio con Emilia Téllez	11,2	24	71.652	STP Santiago S.A.	1°

El día en que se realizó la medición fue el miércoles 03 de Mayo de 13:00 a 17:00. El dispositivo se fijó temporalmente al piso del bus alineado con la dirección del bus para evitar distorsionar los valores de la aceleración. Es importante señalar que según la posición del *smartphone*, cambia la nomenclatura de cada eje de la aceleración. En la Figura 2 se puede apreciar un diagrama con los ejes del dispositivo en la posición usada para medir. Por ejemplo, comúnmente la aceleración en y es la aceleración lateral, pero debido a la posición del equipo, la aceleración en x es la lateral.

Para determinar si la conducción era apropiada o no, se siguió la metodología estipulada en el cálculo del ICA (DTPM, 2013) que utiliza pasajeros incógnitos para la medición de cada atributo. Se incluye además, una serie de criterios de evaluación por cada atributo que buscan reducir la componente de subjetividad que cada pasajero incógnito pueda incorporar. En el caso específico del atributo a02: “El conductor conduce sin frenazos ni movimientos bruscos”, se establecen 5 criterios simples:

- Es considerado frenazo o movimiento brusco cuando un pasajero pierde estabilidad a raíz de la maniobra del conductor.
- Debe disminuir gradualmente la velocidad para pasar por un lomo de toro.

- Debe conducir con la concentración necesaria que le permita detenerse sin frenazos bruscos ante una parada solicitada; luz roja; pare; ceda el paso; paso de cebra y cualquiera tipificada en la ley de tránsito.
- Condiciones de la calzada, virajes estrechos, baches, líneas de tren u otras ajenas al conductor no serían consideradas incumplimiento.
- Si el frenazo o movimiento brusco se realiza con el fin de evitar un accidente, este no será considerado como incumplimiento.



Una vez instalado el dispositivo se comienza la grabación de los datos y cada vez que se genera una maniobra brusca se registra el incidente en el contador y se anota en una bitácora la razón. La aplicación SensorLog se configuró para que almacenara la información con una frecuencia de 60 Hz. Las instancias de conducción brusca se registran por segundo.

De la recolección de datos se pudo armar una base de datos con 235.073 observaciones con una frecuencia aproximada de 60 Hz. La cantidad de atributos relacionados a acelerómetros y giroscopios es 6 (dos por cada eje x, y, z). Mientras que el número de alertas de conducción inapropiada fue 158.

4.3 Procesamiento de datos

Debido a que cada alerta está asociada a un segundo, esto puede llevar fácilmente a errores si es que el pasajero incógnito registró el incidente uno o más segundos después. Para evitar este problema y reducir el número de observaciones de la base de datos, se decide obtener un promedio de los datos del dispositivo inercial cada 3 segundos. Es importante destacar que debido al comportamiento alternante de los datos (ver Figura 3a), un promedio simple cancelaría el efecto que se busca captar.

Por lo tanto, se utilizó el valor cuadrático medio (RMS, por su sigla en inglés) que se define de la siguiente manera:

$$x_{RMS} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N x_i^2} \quad (13)$$

Con esto, los datos de la Figura 3a pasan a verse como en la Figura 3b. Como es de esperar, todos los valores son positivos y se logra capturar los instantes donde la magnitud de la variable aumenta considerablemente.

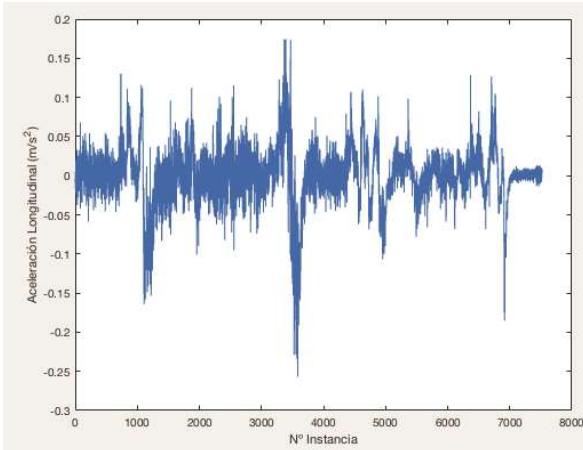


Figura 3a. Aceleración Longitudinal considerando todos los datos reportados por el dispositivo.

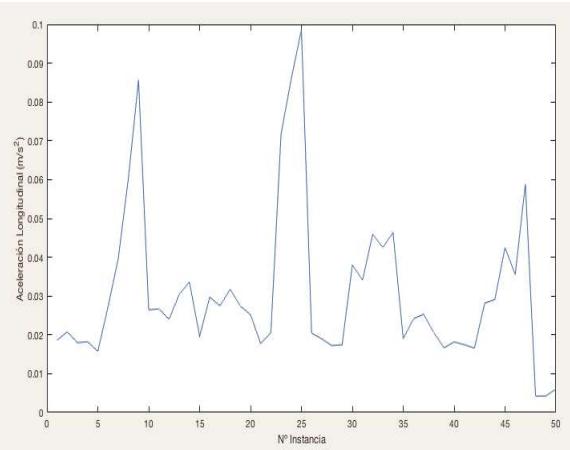


Figura 3b. Aceleración Longitudinal agrupada cada 3 segundos usando el valor cuadrático medio.

En la Figura 4 se grafica la aceleración longitudinal (ya agrupada cada 3 segundos) junto con las alertas registradas por el pasajero incógnito. Es posible notar una gran relación entre los *peaks* de aceleración y la percepción de maniobra de conducción brusca. De todas las variables, *a priori* la aceleración longitudinal es la que mejor explica las alertas. Sin embargo, se consideraron todas las variables para el modelo.

Debido a que los datos deben separarse aleatoriamente entre un set de entrenamiento y un set de testeo, se pierde la secuencialidad de la información. En otras palabras, el dato actual del set (entrenamiento o testeo) ya no tiene relación con el dato anterior ni el siguiente. Pero, esta relación es importante de mantener, ya que cuando se realiza una maniobra brusca, esta es progresiva en el tiempo. Si bien la alerta es en un instante t , la aceleración aumenta desde $t - 1$ y luego se reduce en $t + 1$. Por lo tanto, se crea una nueva base de datos que incorpora esta progresión.

SVM, al igual que muchos modelos de clasificación, funcionan mejor cuando los valores de las variables se encuentran en el rango $[-1,1]$ (Hsu *et al.*, 2016). Es por esta razón que cada vector de atributos (las filas de la matriz) se normaliza. Finalizada esta acción la base está lista para ser transformada en nuestros sets de entrenamiento y testeo.

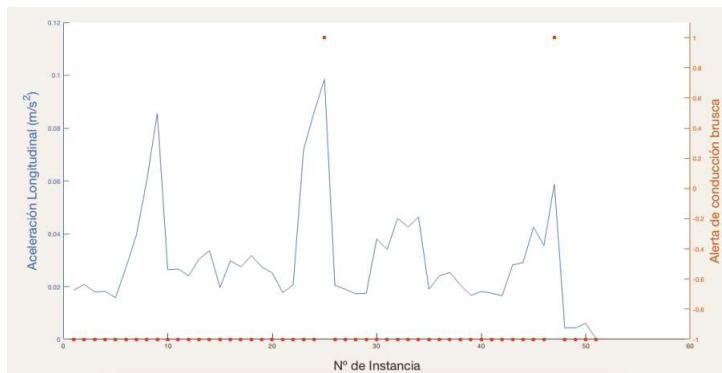


Figura 4. Aceleración Longitudinal con las alertas reportadas (puntos sobre la curva).

5. MODELACIÓN

La proporción de repartición de datos es 80% para el set de entrenamiento y 20% para testeo. Debido a que nuestra muestra sigue siendo no balanceada (a pesar de haber reducido el número de instancias de 235.073 a 1.524), es necesario que la proporción de datos positivos y negativos se mantenga dentro de los sets creados aleatoriamente (en el caso de la base de datos, es 1:13,4). Como primera aproximación, se puede considerar esta proporción como los valores de los parámetros de penalización del error de SVM. Es decir:

$$C_1 = 13,4 \quad ; \quad C_{-1} = 1 \quad (14)$$

Utilizando la librería LIBSVM (Chang y Lin, 2011) en el *software* MATLAB se entrenó el modelo con un kernel RBF y lineal con el objetivo de determinar aquel que entrega mejores rendimientos con los parámetros iniciales. Los resultados se encuentran en la Tabla 4:

Tabla 4. Exactitud del modelo según tipo de Kernel usado.

Tipo de Kernel usado	Exactitud del modelo (10-fold Cross-Validation)
RBF	84,3%
Lineal	85,6%

Debido a lo anterior, se escoge el kernel lineal para entrenar el modelo definitivo. Se deben encontrar los parámetros C_1 y C_{-1} que entreguen mejor rendimiento. Lo usual es tomar en cuenta solo exactitud, pero debido a que la muestra es no balanceada este valor siempre irá disminuyendo. Es por esto que se toman en consideración las otras medidas descritas en la sección 3.4.

A medida que reducimos nuestro error tipo II, el error tipo I y la exactitud del modelo empeoran. Esto lleva a concluir que se debe generar un compromiso entre estos indicadores hasta llegar a los niveles deseados (ver Figura 5). Esta decisión depende completamente de los objetivos que tenga la DTPM o el operador al implementar la metodología. Si se desea estimar muy bien las malas maniobras, hay que tener en cuenta que el costo es aumentar el error tipo I (acusar a un inocente). Es importante notar que el error tipo I cae rápidamente a valores menores de 30%, mientras que el error tipo II solo aumentó a un 5%.

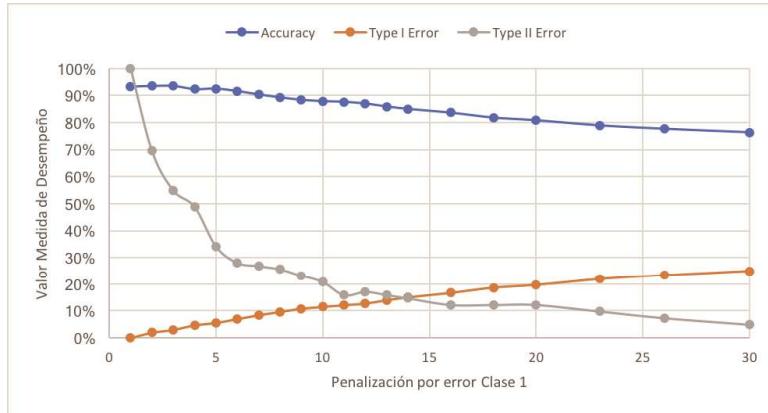


Figura 5. Variación de las medidas de desempeño según el valor del parámetro C_1 .

Como esa decisión escapa los objetivos de este trabajo, se calibrarán dos modelos:

- i) Modelo 1: los errores tipo I y II son similares
- ii) Modelo 2: permite un error tipo II cercano a 30%.

Se generan 15 corridas de entrenamiento/testeo con sets diferentes para evitar obtener resultados poco representativos por la aleatoriedad en la selección de datos para cada set. Para el modelo 1 calibrado se llega a los resultados que se muestran en la Tabla 5:

Tabla 5. Resultados de los modelos calibrados.

Medida	Modelo 1		Modelo 2	
	Promedio	Desv. Est.	Promedio	Desv. Est.
Exactitud	84%	1,6%	89%	1,7%
Error Tipo I	16%	1,9%	9%	2%
Error Tipo II	15%	6,0%	30%	7,7%
Exhaustividad	85%	6,0%	70%	7,7%
Precisión	31%	2,3%	40%	5,4%
F1 Score	45%	2,8%	51%	5,3%

Como era de esperar, el modelo 1 tiene errores tipo I y II de similar valor, cercanos al 15%. La desviación estándar de la medida error tipo II es 6%, la mayor de todas las medidas de desempeño. Esto se debe a que los datos positivos son muy pocos en comparación a los negativos por lo que en cada iteración se produce variabilidad en la predicción. A pesar de ello, se aprecia como en general todas las medidas de desempeño tienen baja desviación entre las 15 iteraciones realizadas.

La precisión del modelo es muy baja comparándola con la exhaustividad (*recall*), esto implica que el modelo encuentra la mayor parte de los datos positivos reales (85%), pero que estos significan apenas el 31% de todos los datos predichos como positivos. Es decir, para capturar todos los culpables, debo atrapar varios inocentes. Si bien esto puede parecer negativo, debido a la naturaleza desbalanceada de los datos recabados, esa proporción de inocentes catalogados como culpables es más bien baja (16%). En números, se clasifican correctamente 21 datos positivos de 25 y 238 datos negativos de 284. Los errores son bajos, pero al comparar tenemos 21

datos TP y 46 datos FP. Por lo tanto, es esperable una precisión tan baja. Como se mencionó anteriormente, esto forma parte del compromiso que debe decidir la autoridad que implemente la metodología.

Al observar los resultados del Modelo 2 (también en la Tabla 4), vemos que el error tipo II es similar a lo buscado (cerca de 30%). Además, el error tipo I es de 9% con una desviación estándar baja, de similar valor, cercanos al 15%. Respecto al Modelo 1, existe mejor exactitud y la precisión del modelo sube 10 puntos porcentuales. Del mismo modo que en el caso anterior, las desviaciones estándar en general son bajas. En números, se clasifican correctamente 18 datos positivos de 25 y 259 datos negativos de 284. Es decir, se perdió 3 TP, pero se ganaron 21 TN. Una muestra más de lo no balanceado de la muestra. Por lo general uno busca tener un menor error tipo I, por lo que el segundo modelo es mejor.

En la Figura 6, se puede observar un gráfico de dispersión entre la aceleración longitudinal y la aceleración lateral, donde los puntos azules son las maniobras reportadas como bruscas. Si bien hay una tendencia a estar separados los positivos de los negativos, existen algunos que parecieran tener las mismas características, pero etiquetados erróneamente.

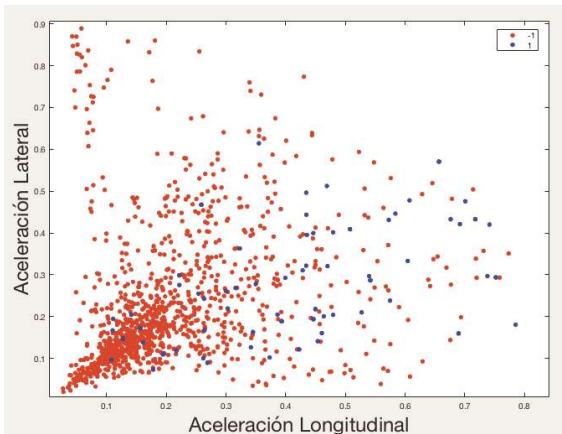


Figura 6. *Scatter Plot* entre aceleración longitudinal y lateral.

Esto se puede deber a:

1. Error humano al momento de etiquetar. La metodología es subjetiva desde el momento que una persona debe juzgar la conducción, si bien las condiciones descritas en la metodología de la DTPM ayudan a reducir la subjetividad, no la eliminan.
2. Dichas condiciones no consideran otras situaciones que pueden llevar a una percepción de mala conducción.
3. Al momento de transcribir las etiquetas se puede estar comentiendo un error debido a la agrupación de 3 segundos.
4. Una maniobra brusca puede tener duración que se escape de los 3 segundos de holgura que se consideró.
5. Pueden existir más variables que expliquen una mala maniobra de conducción, además de las consideradas en el modelo.

Sobre los dos primeros puntos, son problemas que escapan el alcance de esta investigación, ya que se busca utilizar la metodología existente y expandirla usando máquinas de aprendizaje para abarcar toda la flota de buses. Con respecto a los últimos dos puntos, se realizará una revisión de la transcripción de etiquetas, ya que debido al comportamiento de las variables es esperable que los modelos puedan mejorar sus rendimientos con un procedimiento de etiquetado más apropiado.

6. CONCLUSIÓN

Con los resultados que se manejan de los modelos de aprendizaje aplicados se puede concluir que es posible calibrar un SVM con datos pasivos otorgados por dispositivos iniciales. Estos equipos son de bajo costo y están disponibles en todos los celulares modernos, por lo que es factible pensar en una puesta en marcha de esta metodología a mayor escala. Los rendimientos de los modelos dependen bastante del objetivo que se tenga y esta decisión recae en las autoridades competentes en la materia (por ej. DTPM). Si se quiere reducir el error tipo II, pero manteniendo el error tipo I bajo el 10%, el rendimiento del modelo es mejor que si se busca igualar ambos tipos de errores.

Si bien se notó que la predicción no es precisa, esto es fácilmente explicable por la muestra desbalanceada. Además, se cree que la transcripción de etiquetas, desde lo reportado por el pasajero incógnito a la base de datos, puede estar aportando una fuente de error. Esto se revisará en los próximos pasos de la investigación.

A pesar de lo anterior, y en cualquiera de los modelos, la exactitud se mantiene por sobre el 80% y error tipo I por debajo del 15%. Es importante destacar que mientras más datos se le suministren a la máquina de aprendizaje, mejor es la predicción. Si se tiene en consideración que el experimento realizado es más bien reducido y que los resultados son aceptables, una puesta en marcha a gran escala tiene altas probabilidades de resultar exitosa. Además, el esfuerzo logístico y los costos requeridos debieran ser sustancialmente menores a los empleados actualmente.

Como futuros pasos se espera desagregar el rendimiento de los modelos por tramo y comparar la clasificación hecha por SVM con las metodologías encontradas en la literatura que fueron expuestas en el capítulo 2. Es importante señalar que el objetivo de esta investigación es proponer una metodología alternativa que permita su aplicación a todos los buses del sistema, teniendo en consideración que el modelo usado puede no ser el mejor existente.

REFERENCIAS

- Bodini, I., Lancini, M., Pasinetti, S., y Vetturi, D. (2014). Techniques for on-board vibrational passenger comfort monitoring in public transport. *Acta IMEKO*, 3(4), 32-37.
- Boser, B., Guyon, I., y Vapnik, V. (1992). A Training Algorithm for Optimal Margin Classifiers. Artículo presentado en **5th Annual ACM Workshop on Computational Learning Theory**, Pittsburgh, EE.UU.
- Castellanos, J.C., Susin, A., y Fruett, F. (5-7 Octubre 2011). Embedded sensor system and techniques to evaluate the comfort in public transportation. Artículo presentado en **14th International IEEE Conference on Intelligent Transportation Systems**, Washington D.C., EE.UU.

- Chang, C.-C., y Lin, C.-L. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 1-27.
- Chile. Ministerio de Transporte y Telecomunicaciones, Directorio de Transporte Público Metropolitano. (2013). **Manual de Medición ICA 2013**.
- Chile. Ministerio de Transporte y Telecomunicaciones, Directorio de Transporte Público Metropolitano. (2016). **Estudio Satisfacción de Empresas Operadoras**.
- Chile. Ministerio de Transporte y Telecomunicaciones, Directorio de Transporte Público Metropolitano. (2017). **Informe de Gestión 2015-2016**.
- Cortes, C., y Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273-297.
- de Oña, R., Machado, J.L., y de Oña, J. (2015). Perceived Service Quality, Customer Satisfaction, and Behavioral Intentions. *Transportation Research Record: Journal of the Transportation Research Board*, 2538, 76-85.
- dell’Olio, L., Ibeas, A., y Cecin, P. (2011). The quality of service desired by public transport users. *Transport Policy*, 18(1), 217-227.
- Ebrahimi, M.A., Khoshtaghaza, M.H., Minaei, S., y Jamshidi, B. (2017). Vision-based pest detection based on SVM classification method. *Computers and Electronics in Agriculture*, 137, 52-58.
- Figueiredo, M., y Silva, L.F. (12-16 Julio 2015). Public Transport: Whole-Body Vibration and Comfort of Passengers, Drivers and Collectors. Artículo presentado en **22nd International Congress on Sound and Vibration**, Florencia, Italia.
- Grujićić, D., Ivanović, I., Jović, J., y Đorić, V. (2014). Customer perception of service quality in public transport. *Transport*, 29(3), 285-295.
- Han, S. (20-23 Octubre 2013). A study on the ride comfort control using automatic generating data base on board for Electric railway vehicle. Artículo presentado en **13th International Conference on Control, Automation and Systems**, Kimdaejung Convention Center, Gwangju, Corea del Sur.
- Hsu, C.-W., Chang, C.-C., y Lin, C.-J. (2016). **A Practical Guide to Support Vector Classification**.
- Mikrasya (2016). **Counter – Advanced Tally Counter** (Versión 7.1) [Software de Aplicación Móvil].
- Núñez, C., Munizaga, M., y Gschwender, A. (21-25 Octubre 2013). Cálculo de indicadores de calidad de servicio del sistema de transporte público de Santiago a partir de datos pasivos. Artículo presentado en **XVI Congreso Chileno de Ingeniería de Transporte**, Santiago, Chile.
- Ortúzar, J. de D., Willumsen, L. (2011). **Modelling Transport** (4ta. ed.). Chichester, Reino Unido.
- Parasuraman, P., Zeithaml, V., y Berry, L. (1988). SERVQUAL: A multiple – Item Scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40.
- Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J., y White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, 13(4), 295-306.
- Rizzi, L., Cantillo, V., y Bronfman, N. (17-21 Octubre 2005). Influencia de la Percepción del Nivel de Servicio en la Elección Modal para Viajes Interurbanos de Transporte Público hacia el Sur de Chile. Artículo presentado en **XII Congreso Chileno de Ingeniería de Transporte**, Valparaíso, Chile.
- Sánchez-Flores, O., y Romero-Torres, J. (2010). Factores de calidad del servicio en el transporte público de pasajeros: estudio de caso de la ciudad de Toluca, México. *Economía, Sociedad y Territorio*, 10(32), 49-80.
- Thomas, B. (2017). **SensorLog** (Versión 1.9.2) [Software de Aplicación Móvil].
- Wang, X. (2015). Short Term Prediction of Freeway Exiting Volume Based on SVM and KNN. *International Journal of Transportation Science and Technology*, 4(3), 337-352.
- Zhang, K., Zhou, K., y Zhang, F. (2014). Evaluating bus transit performance of chinese cities: Developing an overall bus comfort model. *Transportation Research Part A: Practice and Policy*, 69, 105-112.