

UNIVERSIDAD MAYOR DE SAN SIMÓN

PROYECTO DE GRADO

Detección de anomalías de conducción

Autor:
Evelyn CUSI LÓPEZ

Tutor:
Dr. Eduardo DI SANTI

*Un proyecto de grado presentado en cumplimiento de los requisitos
para el título de Licenciatura en Ingeniería de Sistemas*

en la

Carrera de Ingeniería de Sistemas
Departamento Informática-Sistemas

8 de septiembre de 2018

Agradecimientos

Le agradezco a Dios por haberme acompañado y guiado a lo largo de mi carrera, por ser mi fortaleza en los momentos de debilidad y por brindarme una vida llena de aprendizajes, experiencias y sobre todo felicidad.

A mi madre, por estar conmigo, por enseñarme a crecer y a levantarme, por apoyarme y guiarme, por ser la base que me ayudó a llegar hasta aquí.

Índice general

Agradecimientos	III
Resumen	IX
1. Introducción	1
1.1. Planteamiento del problema	1
1.2. Objetivo general	3
1.3. Objetivos específicos	3
1.4. Justificación	3
1.4.1. Justificación práctica	3
1.4.2. Justificación metodológica	3
1.4.3. Justificación teórica	4
1.5. Límites y alcances	4
1.6. Método de investigación	4
2. Fundamentos de la detección de anomalías	5
2.1. Detección de anomalías	5
2.1.1. Enfoques de detección de anomalías	7
Detección de anomalías supervisada	7
Detección de anomalías semi-supervisada	8
Detección de anomalías no supervisada	8
2.2. Trabajo relacionado	8
2.3. Enfoque sobre el problema	9
3. Captura y análisis de datos	11
3.1. Captura de datos	11
3.2. Análisis de los datos	12
3.2.1. Normalización	14
Escalado de variables (Feature Scaling o MinMax Scaler)	15
Escalado estándar (Standard Scaler)	15
Escalado sobre el valor máximo	15
Escalado robusto (Robust scaler)	15

UNIVERSIDAD MAYOR DE SAN SIMÓN

Resumen

Facultad de Ciencias y Tecnología
Departamento Informática-Sistemas

Licenciatura en Ingeniería de Sistemas

Detección de anomalías de conducción

por Evelyn CUSI LÓPEZ

El presente trabajo describe el desarrollo de un mecanismo de detección de anomalías de manejo, el cual es implementado usando un dispositivo móvil y técnicas de Aprendizaje Automático.

El objetivo es crear una herramienta capaz de encontrar comportamientos anómalos en el manejo de un usuario concreto, teniendo un previo conocimiento de las conductas normales de manejo del mismo. Asimismo se presenta antecedentes de trabajos e investigaciones de la detección de anomalías de manejo de todo el mundo, se analiza los parámetros de manejo de conducción obtenidos por el dispositivo móvil, y se presenta la propuesta para identificar anomalías mediante Redes Neuronales Recurrentes, un método de Aprendizaje Automático que es comunmente utilizado para series de tiempo.

El trabajo cuenta con dos partes principales: un modelo ajustado al comportamiento normal de manejo de un usuario, y un método de detección de anomalías.....

La precisión de detección del método proupuesto en este proyecto es de que fue evaluado con

Capítulo 1

Introducción

El presente documento describe el desarrollo de un método para la detección de anomalías en la conducción. Se propone el uso de técnicas de Aprendizaje Automático para generar un mecanismo que identifique anomalías de manejo, de tal modo que éstas puedan usarse para alertar oportunamente a los conductores y así logren corregir sus conductas de manejo.

La idea principal del presente trabajo consiste en aprender el comportamiento normal de conducción, para posteriormente detectar de forma autónoma aquellos comportamientos inesperados e informarlos como anomalías, de manera que se pueda evitar un accidente de tránsito o reducir los efectos del mismo.

1.1. Planteamiento del problema

Debido a las graves secuelas que causan sobre las personas y los altos costos económicos asociados a ellos, los accidentes de tránsito se catalogan como un problema social y de salud pública mundial.

Según la Organización Mundial de Salud (OMS) cada año existen aproximadamente 1,25 millones de muertes a causa de accidentes de tránsito, agregando que la mitad de todas estas víctimas son peatones, ciclistas y motociclistas (Véase la figura 1.1 pag. 2). Asimismo se puede decir que son una de las causas de muerte más importantes en el mundo, y la principal causa de muerte entre personas de edades comprendidas entre los 15 y los 29 años.

Los accidentes registrados en 2017, según la Unidad Operativa de Tránsito de Cochabamba, provocaron la muerte de 200 personas y dejaron aproximadamente 2200 heridos.

En la figura 1.2 se muestra las causas por las cuales se ocasiona un accidente de tránsito, se puede observar que gran parte de éstas se deben al factor humano sin embargo hay otras que conllevan factores medio-ambientales y mecánicos, por lo que se hace imposible evitar completamente los mismos.

Es por ello que se hace necesario el contar con mecanismos para prevenir y/o actuar de forma oportuna ante posibles accidentes de tránsito, motivo por el cual el presente trabajo se centra en estudiar los comportamientos de conducción, para así generar

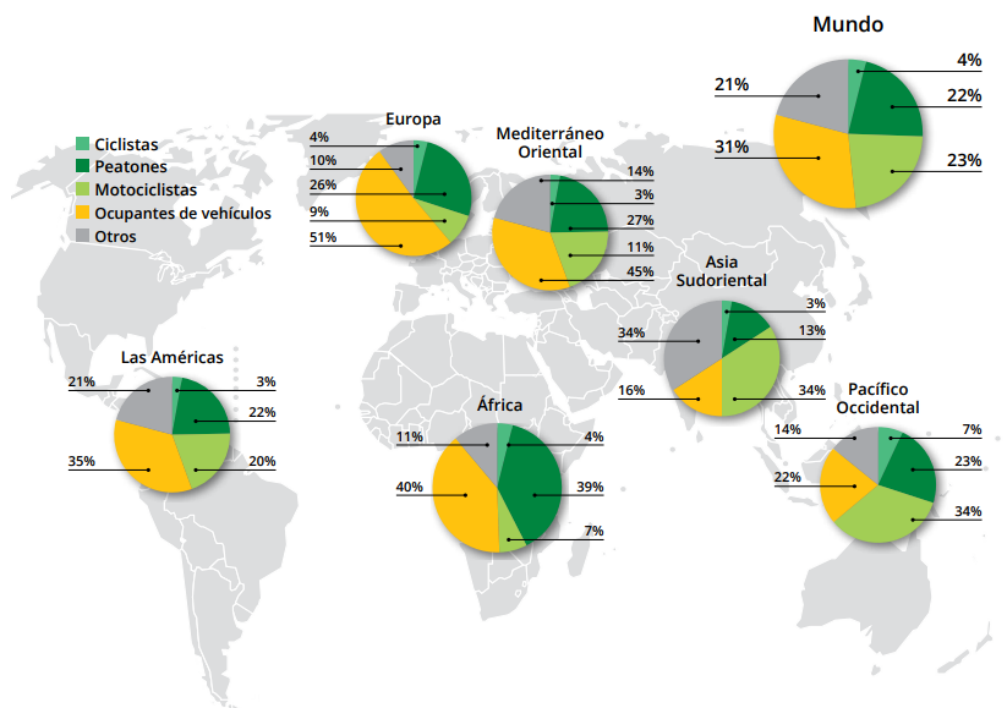


FIGURA 1.1: Muertes por accidentes de tránsito por región en función del tipo de usuario (2013), OMS

alertas al encontrar un comportamiento anómalo en el manejo, de manera que se pueda evitar o en todo caso minimizar los efectos del mismo.

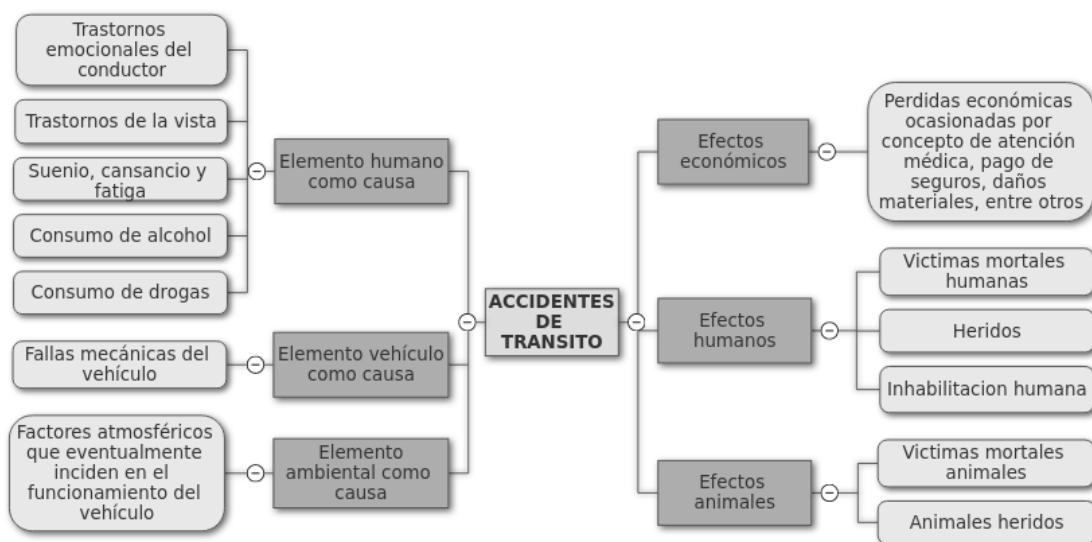


FIGURA 1.2: Árbol de problemas

1.2. **Objetivo general**

El objetivo general del presente trabajo es desarrollar un mecanismo de detección de anomalías de conducción, mediante el uso de un dispositivo móvil y algoritmos de Aprendizaje Automático, con el fin de alertar de forma oportuna el hallazgo de un patrón anómalo en el manejo, tal como cansancio, ebriedad, o problemas de salud, ej, epilepsia.

1.3. **Objetivos específicos**

- Capturar los parámetros de manejo de un conductor mediante el uso de sensores de un dispositivo móvil.
- Escalar los parámetros de manejo mediante técnicas de pre-procesamiento de datos.
- Generar un modelo de aprendizaje automático que se ajuste a un comportamiento normal de manejo.
- Definir un método de detección de anomalías para generar una alerta de conducción anormal.
- Evaluar el método de detección de anomalías con nuevas muestras.

1.4. **Justificación**

Los accidentes de tránsito cobran un número inaceptable de víctimas cada año, especialmente en las regiones mas pobres del mundo. Esto se debe a diversos aspectos, pero el principal recae en el bajo nivel de conciencia ciudadana que existe, lo que conlleva a que muchas personas manejen bajo los efectos del alcohol, con exceso de velocidad, manipulando sus dispositivos móviles, entre otros. Por ello este trabajo busca establecer patrones de comportamientos de conducción mediante el uso de un dispositivo móvil y técnicas de Aprendizaje Automático, de manera que se logre realizar una detección de anomalías oportuna.

1.4.1. **Justificación práctica**

Detectar anomalías de conducción permite generar una alerta oportuna al conductor para que éste logre corregir sus conductas de manejo de forma rápida. De esta manera se podrá evitar accidentes de tránsito o minimizar los efectos del mismo, permitiendo así reducir la cantidad de daños, tanto materiales como personales.

1.4.2. **Justificación metodológica**

El estudio realizado en el desarrollo del presente proyecto de grado permite resaltar la eficiencia de las técnicas de Aprendizaje Automático en la detección de anomalías.

1.4.3. Justificación teórica

1.5. Límites y alcances

Debido a que la realización de pruebas de campo de ésta investigación es bastante peligroso, se limitó los ejemplos de conducción anómala a:

Siendo así los experimentos y pruebas se realizaron sólo sobre un pequeño conjunto de ejemplos anómalos, por lo tanto no se espera que éste funcione de manera correcta sobre aquellos ejemplos que no fueron considerados.

1.6. Método de investigación

La presente investigación se realizó con un enfoque experimental, teniendo como hipótesis la siguiente:

¿Es posible detectar anomalías de conducción mediante el uso de un dispositivo móvil y algoritmos de Aprendizaje Automático?

Capítulo 2

Fundamentos de la detección de anomalías

En este capítulo se aborda los conceptos necesarios que se necesitan para comprender la detección de anomalías, así también se expondrá los diferentes proyectos e investigaciones enfocados en la detección de anomalías de conducción.

2.1. Detección de anomalías

Las anomalías, o valores atípicos, son patrones en los datos que no se ajustan a una noción bien definida de un comportamiento normal, por lo tanto, la detección de tales patrones en los datos, que son diferentes del comportamiento normal esperado, es lo que se conoce como **detección de anomalías**.

La detección de anomalías, también conocida como detección de valores atípicos, se usa en distintos dominios de aplicaciones, por ejemplo: procesamiento de imágenes, detección de fraudes de tarjetas, sistemas de detección de intrusiones de red y muchos otros.

Las anomalías pueden ser clasificadas dentro de una de las tres siguientes categorías:

1. **Anomalías de punto:** Las anomalías de punto son simplemente instancias únicas y anómalas dentro de un conjunto de datos más grande, es decir, éstas se encuentran separadas del resto de los datos. Por ejemplo, en la Figura 2.1, los puntos o_1 , o_2 y la región O_3 se encuentran fuera de los límites de las regiones normales (N_1 y N_2), y por lo tanto son anomalías puntuales ya que son diferentes a los puntos de datos normales.

Este es el tipo más simple de anomalía y es el foco de la mayoría de las investigaciones enfocados en la detección de valores atípicos..

2. **Anomalías contextuales (o condicionales):** Estos son puntos que solo se consideran anómalos en un contexto específico. La noción de este contexto es inducida por la estructura en el conjunto de datos y debe especificarse como parte de la formulación del problema.

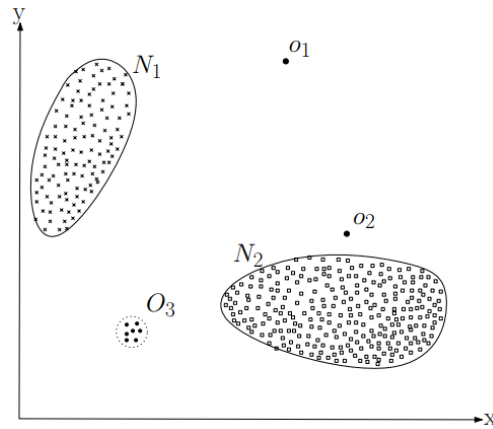


FIGURA 2.1: Ejemplo de anomalía en con conjunto de datos de 2 dimensiones.

Este tipo de anomalías se han explorado con mayor frecuencia en los datos de series de tiempo y datos espaciales. La figura 2.2 muestra un ejemplo de serie temporal de la temperatura mensual de un área en los últimos 5 años, se debe tener en cuenta que la temperatura en el tiempo t_1 es la misma que en el tiempo t_2 , pero se produce en un contexto diferente, por lo tanto t_2 es considerada una anomalía.

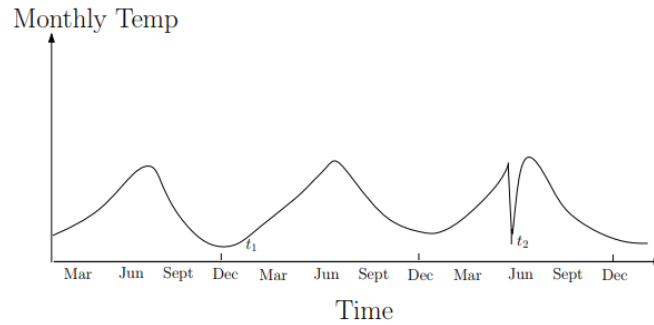


FIGURA 2.2: Anomalía contextual t_2 en una serie temporal de temperatura.

3. **Anomalías colectivas:** Si una colección de instancias de datos relacionadas es anómala con respecto a todo el conjunto de datos, se denomina anomalía colectiva. Las instancias de datos individuales en una anomalía colectiva pueden no ser anomalías por sí mismas, pero su aparición conjunta como una colección es anómala.

La figura 2.3 ilustra un ejemplo que muestra una salida de electrocardiograma humano, se puede notar que la región resaltada denota una anomalía porque existe el mismo valor bajo durante un tiempo anormalmente prolongado (que corresponde a una Contracción prematura auricular). Se debe tener en cuenta que ese valor bajo por sí mismo no es una anomalía.

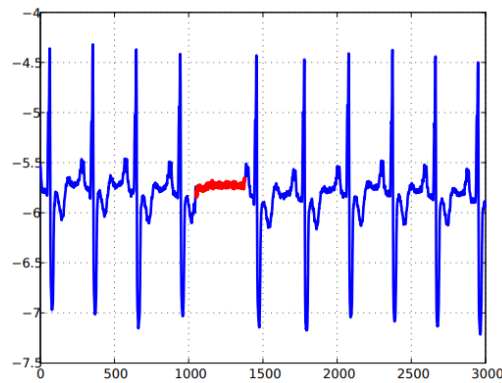


FIGURA 2.3: Anomalía colectiva correspondiente a una contracción prematura auricular en un electrocardiograma humano.

En un nivel abstracto, la detección de anomalías puede parecer una tarea simple. Sin embargo puede llegar a ser una tarea muy desafiante. A continuación se presenta algunos de estos desafíos.

- La definición de regiones normales es bastante difícil. En muchos casos, los límites entre las anomalías y los datos normales no son precisos. Por lo tanto, las observaciones normales podrían considerarse anomalías y viceversa.
- Lo que se considera normal hoy en día, puede no ser normal en el futuro.
- La mayor parte de las veces los enfoques para la detección de anomalías en un campo en específico no se pueden utilizar en otro campo.
- La poca disponibilidad de datos de capacitación y validación para el entrenamiento del modelo.

2.1.1. Enfoques de detección de anomalías

Los enfoques que se pueden usar para encontrar anomalías se clasifican en las siguientes categorías:

Detección de anomalías supervisada

El uso de técnicas de aprendizaje supervisado requiere la disponibilidad de un conjunto de datos de entrenamiento etiquetados, tanto para clases normales como anómalas. El enfoque principal es construir un modelo predictivo para clases normales vs. anomalías, posteriormente tomar cualquier instancia de datos no visto, comparar con el modelo y determinar a que clase pertenece.

Existen dos principales inconvenientes que surgen con el uso de ésta técnica.

- La cantidad de instancias anómalas es muy inferior a la de las instancias normales, lo que genera un desequilibrio de distribución de clases durante el entrenamiento.

- La obtención de etiquetas precisas y representativas, en particular para la clase de anomalía, es un desafío.

Detección de anomalías semi-supervisada

Estas técnicas requieren un conjunto de entrenamiento con instancias etiquetadas, pero solo para la clase normal, esto hace que su uso sea más aplicable que las técnicas supervisadas, ya que no se requiere etiquetas para la clase anomalía.

El enfoque típico usado en éstas técnicas es construir un modelo para la clase correspondiente al comportamiento normal, y usar el modelo para identificar anomalías en los datos de prueba.

Detección de anomalías no supervisada

Las técnicas que operan de manera no supervisada no requieren datos de entrenamiento, razón por la cual son las más ampliamente utilizadas. Estas técnicas suponen que las instancias normales son mucho más frecuentes que las anomalías en los datos de prueba, en caso de que esta suposición no sea cierta, tales técnicas sufren de una alta tasa de falsas alarmas.

2.2. Trabajo relacionado

La identificación de comportamientos de conducción anormal es una parte indispensable para mejorar la seguridad de conducción, sin embargo, como se vió previamente, ésta no es una tarea sencilla. En los últimos años, se han propuesto varias técnicas para detectar conductas de manejo. Esta sección esta dedicada al repaso de las mismas.

Dang-Nhac et al. [20], proponen un sistema combinado que se compone de dos módulos: uno para detectar el tipo de vehículo de los usuarios y el otro para detectar los eventos de conducción instantánea, independientemente de la orientación y la posición de los teléfonos inteligentes, éste sistema logra una precisión promedio del 98.33 % en la detección del tipo del vehículo (automóvil, motocicleta, bicicleta, entre otros) y una precisión promedio de 98.95 % en el reconocimiento de los eventos de conducción de los motociclistas al usar Random Forest como clasificador .

Por otra parte en [21] Ferreira et al. presentan una evaluación cuantitativa de 4 Algoritmos de Aprendizaje Automático (Bayesian Network BN, Artificial Neural Network ANN, Random Forest RF y Support Vector Machine SVM) con diferentes configuraciones, aplicadas en la detección de 7 tipos de eventos de conducción, entre eventos normales y agresivos, utilizando datos recopilados de 4 sensores de teléfonos inteligentes Android (acelerómetro, aceleración lineal, magnetómetro y giroscopio); dando como resultado que el giroscopio y el acelerómetro son los mejores sensores para detectar eventos de conducción y que Random Forest (RF) es por lejos el Algoritmo de Aprendizaje Automático de mejor rendimiento, seguido de la forma más simple de ANN el Multi Layer Perceptron (MLP).

En [23] proponen el sistema MIROAD el cual muestra que el Dynamic Time Warping (DTW) es un algoritmo válido para detectar maniobras de conducción potencialmente agresivas, donde casi todos los eventos agresivos (97 %) se identificaron correctamente, utilizando el conjunto de sensores T (acelerómetro, giroscopio y el tono de voz). Así también en [24] se propone un sistema enfocado a desarrollar la conciencia del conductor mediante notificaciones en situaciones críticas que pueden desencadenar maniobras de manejo inseguras, mediante un modelo para detectar situaciones peligrosas basadas en Object-oriented Bayesian Network (OOBN).

Así como los trabajos presentados previamente existe gran cantidad de trabajos ([25], [26], [27], [28], [29]) que usan los sensores de los teléfonos inteligentes (acelerómetro y giroscopio) para la detección de conducción agresiva, esto debido a que se tiene la ventaja de no comprar ni instalar ningún dispositivo y además de ser altamente portátil, sin embargo se depende bastante del rendimiento del receptor GPS y no es aplicable en áreas no disponibles para GPS.

Existen además otros enfoques para la detección de conducción agresiva de un conductor, por ejemplo en [22] proponen un método basado en Convolutional Neural Network (CNN) para detectar la emoción de conducción agresiva, mediante la utilización de imágenes faciales de un conductor obtenidas con una cámara de luz NIR y una cámara térmica.

2.3. Enfoque sobre el problema

Es evidente que éste tema fue ampliamente investigado y que tiene una gran variedad de propuestas de solución, sin embargo la mayoría de estas se basan en la detección mediante técnicas de aprendizaje supervisado, lo cual presenta la gran desventaja de requerir datos etiquetados para generar el modelo de detección; además gran parte de los trabajos relacionados proponen modelos generalizados para la detección y no así modelos específicos por usuarios, lo cual es crucial debido a que cada conductor presenta conductas individuales de manejo y manejan en condiciones distintas, es decir, la conducción de un usuario que circula por avenidas pavimentadas será distinta a la conducción de un usuario que circula por calles empedradas o la conducción de un usuario que circula por avenidas o calles concurridas será distinta a de los usuarios que circulen por calles relativamente descongestionadas.

Es por ello que el presente proyecto propone un método de detección de anomalías de conducción (Figura 2.4) como alternativa a las ya existentes.

Este método se compone de las siguientes partes principales:

- Captura y análisis de datos.
- Generación del modelo de conducta normal de conducción.
- Detector de anomalías.

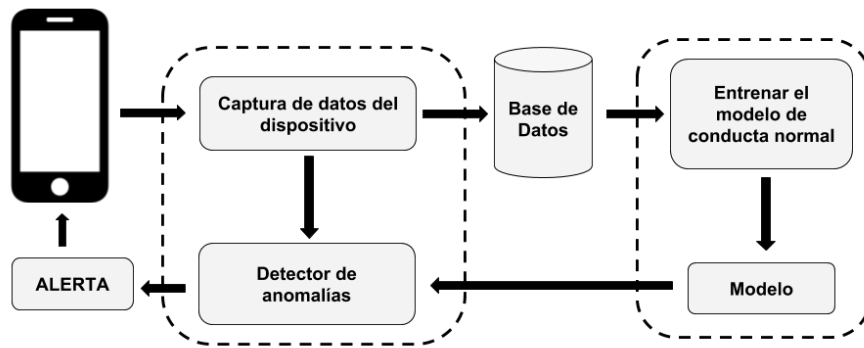


FIGURA 2.4: Método de detección propuesto.

Los cuales serán abordados en los siguientes capítulos.

Capítulo 3

Captura y análisis de datos

Contar con una gran cantidad de datos en cualquier problema de detección de anomalías es lo que permite generar modelos más precisos, debido a que nunca se sabe qué características pueden dar indicio de una anomalía, contar con múltiples tipos de datos es lo que permite ir más allá de una mera detección de anomalías puntuales y ser capaz de identificar anomalías contextuales o colectivas más sofisticadas. Sin embargo la obtención de estos datos no siempre es una tarea sencilla, por lo que muchas veces se debe encontrar una manera de generar los mismos.

Este capítulo detalla el método de recolección de datos que se realizará para la presente investigación y las diferentes técnicas de análisis de datos que se aplicará.

3.1. Captura de datos

Actualmente existen varios enfoques para acceder a la información del conductor y del vehículo. En el primer enfoque, un conjunto de sensores y hardware adicional se implementan previamente en el vehículo, por ejemplo, cajas telemáticas (cajas negras provistas por compañías de seguros de automóviles), adaptadores de diagnóstico a bordo (OBD-II) enchufados en el controlador del vehículo red de área (CAN) ([30], [31]), la información registrada por estos dispositivos se puede recuperar o enviar a través de Internet. Sin embargo, esta estrategia requiere que los vehículos instalen dispositivos adicionales, lo que implica un mayor costo. Para superar estos inconvenientes, existe un enfoque alternativo el cual es usar teléfonos inteligentes para recopilar datos a través de un conjunto de sensores integrados, tales como sensores inerciales (acelerómetros y giroscopios), sistemas de posicionamiento global (GPS), magnetómetros, micrófonos, sensores de imagen (cámaras), sensores de luz, sensores de proximidad, sensores de dirección (brújula), entre otros.

Para el presente proyecto se eligió el uso de teléfonos inteligentes para acceder a la información del tipo de conducción, por las razones que se presentaron anteriormente, con este enfoque se desarrolló una aplicación móvil basada en Android para recopilar datos de los sensores: acelerómetro y giroscopio, en intervalos de 1 segundo, los cuales en una primera instancia serán almacenados de manera interna en el dispositivo móvil. (Ver Figura 3.1)

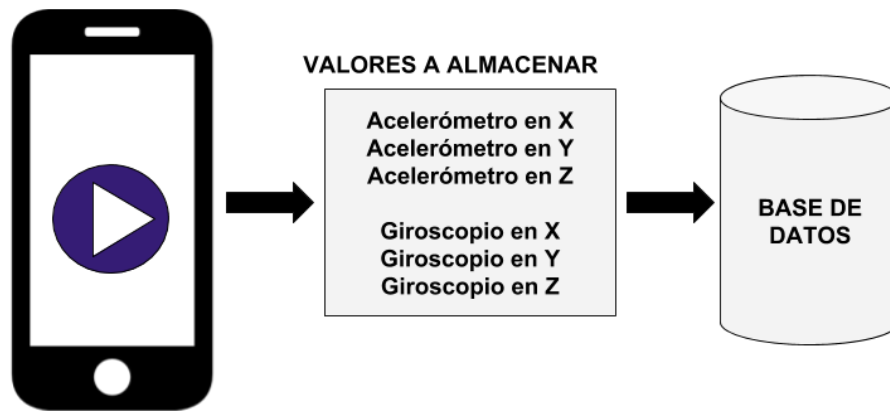


FIGURA 3.1: Recolección de datos, con intervalo de un segundo.

Para la captura de datos se usó un soporte para celular de parabrisas como se ve en la Figura 3.2; se realizó la captura en dos posiciones distintas (vertical y horizontal).



FIGURA 3.2: Soporte para celular de parabrisas, posición horizontal.

Cada captura, independientemente de la posición en la que se realizó, dió como resultado un conjunto de datos (dataset), donde por cada tiempo T (1 seg.) se tiene seis variables: acelerómetro en X (acc x), acelerómetro en Y (acc y), acelerómetro en Z (acc z), giroscopio en X (gyr x), giroscopio en Y (gyr y) y giroscopio en Z (gyr z). En la Figura 3.3 se aprecia un fragmento del conjunto de datos que se obtuvo en una captura.

3.2. Análisis de los datos

Como se indicó en la anterior sección, se realizó la captura de datos del manejo de un usuario (conductor), de forma vertical y horizontal, a continuación se analizará las diferencias entre ellos.

_id	acc_x	acc_y	acc_z	gyr_x	gyr_y	gyr_z
	Filter	Filter	Filter	Filter	Filter	Filter
1	-0.4353179...	9.60725402...	1.84175109...	9.15527343...	-0.0020294...	0.00770568...
2	-1.1603393...	9.73767089...	1.84773254...	-0.0500488...	0.02827453...	0.00379943...
3	-0.3013153...	9.57853698...	1.90994262...	0.00070190...	-0.0008697...	0.00354003...
4	-0.3479614...	9.59648132...	1.89201354...	0.00035095...	-0.0009918...	0.00372314...

FIGURA 3.3: Fragmento del conjunto de datos obtenido.

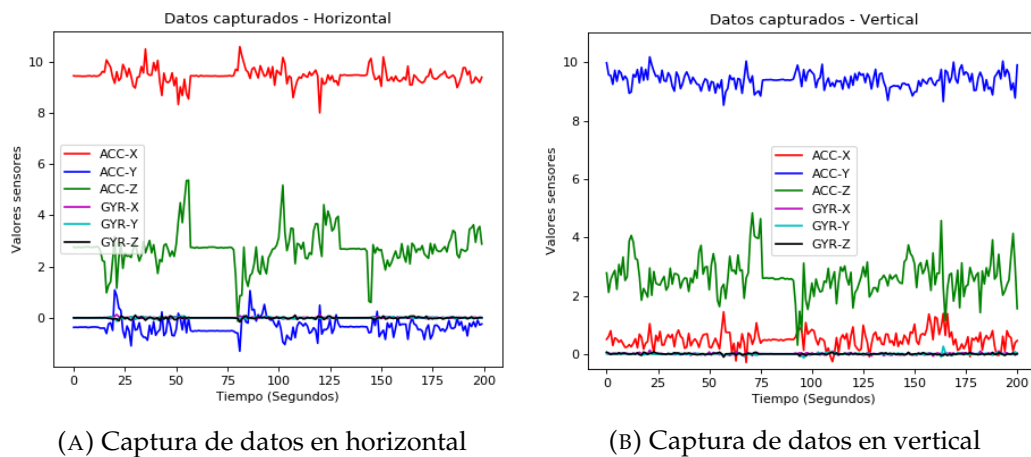


FIGURA 3.4: Grafica de los sensores capturados en diferentes posiciones

En la Figura ?? se muestra fragmentos de las capturas obtenidas por el dispositivo móvil del manejo de un mismo usuario desde diferentes posiciones.

Si bien los valores capturados, son muy similares entre sí, los datos que fueron capturados con el dispositivo móvil en posición horizontal, presentan menos ruido, esto debido a que esta posición favorece la inercia del dispositivo cuando el vehículo está en movimiento, lo cual es una gran ventaja frente a los datos que fueron capturados de forma vertical, ya que estos fueron más susceptibles a sacudirse mientras el vehículo se desplazaba haciendo que los valores capturados en esta posición presenten valores de movimiento no sólo del vehículo sino también del dispositivo móvil, lo cual no es lo que se busca en el presente proyecto.

Por las razones presentadas en el anterior párrafo se decidió trabajar con los datos capturados con el dispositivo móvil en posición horizontal.

Para obtener un análisis más profundo de los datos obtenidos, a continuación se presenta una tabla con los valores estadísticos de cada columna de datos capturada (Tabla 3.1).

	ACC-X	ACC-Y	ACC-Z	GYR-X	GYR-Y	GYR-Z
Min	-16.1556	-11.5167	-11.2487	-3.0617	-3.2679	-4.7595
Max	15.5559	11.3349	15.2552	4.3997	2.3482	2.9745
Media	9.5361	-0.7939	-0.4447	0.0015	0.0003	-0.0011
Mediana	9.7751	-0.9049	-0.4820	-0.0001	6.1035e-05	6.1035e-05
Des. est.	1.7220	1.3904	0.9007	0.0905	0.0857	0.1042

CUADRO 3.1: Tabla de resultados estadísticos del conjunto de datos.

Con la información proporcionada por la tabla ahora es más fácil realizar el análisis del conjunto de datos, en primer lugar se evidencia que los valores obtenidos durante la captura, con el uso del dispositivo móvil, se presentan entre rangos de valores muy distintos, se puede observar por ejemplo, que el valor del acelerómetro en X oscila entre los valores de -16 y 15 aproximadamente, mientras que los valores del giroscopio en X se encuentran entre -3 y 4, lo cual representa un problema debido a que ésta diferencia de rangos puede generar más peso en unos datos que en otros.

Para notar de una manera más gráfica el problema expuesto anteriormente a continuación se presenta histogramas de las frecuencias de valores por cada tipo de valores que existe en el conjunto de datos. (Ver Figura 3.5)

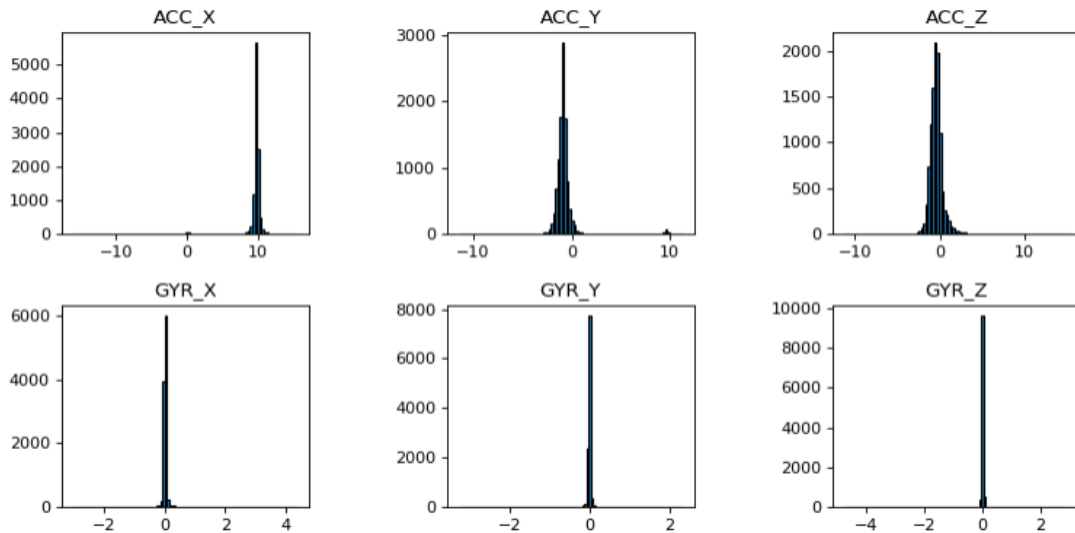


FIGURA 3.5: Histograma de frecuencias del conjunto de datos

Para solucionar este problema existe diversas técnicas, los cuales se abordarán en la siguiente sección.

3.2.1. Normalización

La normalización de datos es un proceso o técnica para comprimir o extender los valores de una variable para que estén en un rango definido, lo cual hace que muchos

algoritmos de Aprendizaje Automático funcionen mejor, sin embargo, una mala aplicación de este proceso, o una elección descuidada del método de normalización puede arruinar los datos.

A continuación se presenta algunos ejemplos de los métodos de normalización más utilizados.

Escalado de variables (Feature Scaling o MinMax Scaler)

En este método, cada entrada se normaliza entre unos límites definidos:

$$x_{normalizado} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.1)$$

Presenta el problema de que comprime los datos de entrada entre unos límites fijos, que por lo general son 0 y 1. Esto quiere decir que si existe ruido, éste va a ser ampliado, lo que hace que este método no sea adecuado para señales estables.

Escalado estándar (Standard Scaler)

Es un método alternativo al escalado de variables, consiste en restar a cada dato la media de la variable y dividirlo por la desviación típica.

$$x_{normalizado} = \frac{x - x_{media}}{x_{desvSt}} \quad (3.2)$$

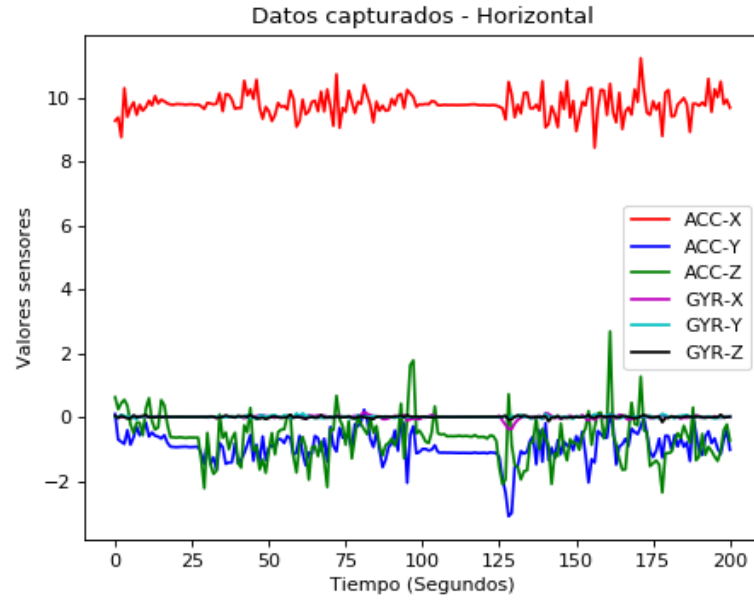
Éste método si sería adecuado para normalizar señales estables, no obstante, tanto la media como la desviación estándar son muy sensibles a valores anómalos. Una alternativa de solución de esto sería la eliminación de anomalías antes de realizar la normalización.

Escalado sobre el valor máximo

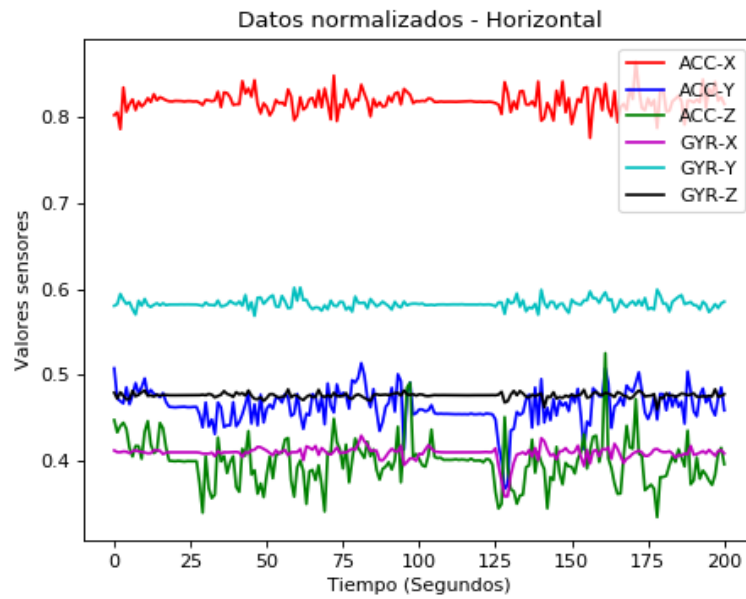
El siguiente método, presenta la idea de escalar los datos dividiendo éstos entre su máximo valor.

Escalado robusto (Robust scaler)

El escalado robusto consiste en eliminar la mediana y escala los datos de acuerdo con el rango de intercuartil (IQR). Este método es robusto para valores atípicos.



(A) Datos puros



(B) Datos normalizados

FIGURA 3.6: Gráfica de de comparación entre datos puros y normalizados

Bibliografía

- [1] DANG-NHAC, L., DUC-NHAN, N., THI-HAU, N., AND HA-NAM, N. Vehicle mode and driving activity detection based on analyzing sensor data of smartphones, *Sensors* v. 18, 4 (2018), 1036.
- [2] FERREIRA, J., CARVALHO, E., FERREIRA, B., DE SOUZA, C., SUHARA, Y., PENTLAND, A., AND PESSIN, G. Driver behavior profiling: An investigation with different smartphone sensors and machine learning, *PLoS One* v. 12, 4 (2017), e0174959.
- [3] WHO-LEE, K., SIK-YOON, H., MIN-SONG, J., AND RYOUNG-PARK, K. Convolutional Neural Network-Based Classification of Driver's Emotion during Aggressive and Smooth Driving Using Multi-Modal Camera Sensors, *Sensors* v. 18, 4 (2018), 957.
- [4] JOHNSON, D., AND TRIVEDI, M. Driving Style Recognition Using a Smartphone as a Sensor Platform. En: *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference*, 2011. p. 1609–1615.
- [5] KRIDALUKMANA, R., YAN-LU, H., AND NADERPOUR, M. An Object Oriented Bayesian Network Approach for Unsafe Driving Maneuvers Prevention System, 2017 12th International IEEE Conference.
- [6] BHOYAR, V., LATA, P., KATKAR, J., PATIL, A., AND JAVALE, D. Symbian Based Rash Driving Detection System. *Int. J. Emerg. Trends Technol. Comput. Sci.* 2013;2:124–126.
- [7] CHEN, Z., YU, J., ZHU, Y., CHEN, Y., AND LI, M. D3: Abnormal Driving Behaviors Detection and Identification Using Smartphone Sensors; *Proceedings of the 12th Annual IEEE International Conference on Sensing, Communication, and Networking*; Seattle, WA, USA. 22–25 de Junio 2015; pp. 524–532.
- [8] EREN, H., MAKINIST, S., AKIN, E., AND YILMAZ, A. Estimating Driving Behavior by a Smartphone; *Proceedings of the Intelligent Vehicles Symposium*; Alcalá de Henares, España. 3–7 de Junio 2012; pp. 234–239.
- [9] BOONMEE, S., AND TANGAMCHIT, P. Portable Reckless Driving Detection System; *Proceedings of the 6th IEEE International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*; Pattaya, Chonburi, Tailandia. 6–9 de Mayo 2009; pp. 412–415.
- [10] KOH, D.-W., AND KANG, H.-B. Smartphone-Based Modeling and Detection of Aggressiveness Reactions in Senior Drivers; *Proceedings of the IEEE Intelligent Vehicles Symposium*; Seoul, Korea. 28 de Junio – 1 de Julio 2015; pp. 12–17.

- [11] ZALDIVAR, J., CALAFATE, C.T., CANO, J.C., AND MANZONI, P. Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones; Proceedings of the 2011 IEEE 36th Conference on Local Computer Networks; Bonn, Alemania. 4–7 de Octubre 2011; pp. 813–819.
- [12] ARAUJO, R., IGREJA, A., DE CASTRO R., AND ARAUJO, R. Driving coach: A smartphone application to evaluate driving efficient patterns; Proceedings of the 2012 IEEE Intelligent Vehicles Symposium (IV); Alcala de Henares, España. 3–7 de Junio 2012; pp. 1005–1010.