

PREDICCIÓN DEL ESTADO DEL TRANSITO UTILIZANDO APRENDIZAJE DE MÁQUINA

MSc. VÁZQUEZ Raimundo; FERNÁNDEZ, Juan M.; SIMÓN, Francisco;
Ing. OLIVETTI, Alejandra (*)

RESUMEN:

Se desarrolla un procedimiento para predecir el estado del tránsito de un sistema vehicular utilizando como herramientas: dispositivos genéricos, simulación de un sensor laser y un Aprendizaje de Máquina. Tal predicción se realiza mediante tres modelos: Clasificador Bayer-Navier, Árbol de decisiones y Redes neuronales. Para ello fue necesario realizar una etapa de entrenamiento previa a la ejecución de cada algoritmo. Luego un programa de alto nivel articulado con la librería OpenCV se encarga de fusionar los datos obtenidos del ambiente vehicular y realizar la predicción.

Palabras clave: Aprendizaje de Máquina, Clasificadores Bayesianos, Redes neuronales, Árbol de decisión

INTRODUCCIÓN

El propósito de este trabajo es desarrollar un procedimiento para predecir el estado del tránsito empleando Inteligencia Artificial.

De esta manera se pretende contribuir en soluciones tecnológicas para abordar los problemas de seguridad vial.

HERRAMIENTAS UTILIZADAS

Aprendizaje de Máquina

Machine Learning (ML) es una disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente (Hastie T., Tibshirani R., Friedman J., 2001; Bradski G., Kaebler A., 2008; Duda R. O., Hart P. E., 1973). Aprender en este contexto quiere decir identificar patrones complejos entre muchos datos. El concepto de ML es un algoritmo que revisa los datos y es capaz de predecir comportamientos futuros en forma automática utilizando un sistema de clasificación. Estos sistemas tienen como función predecir una categoría, que en nuestro caso se asocia a un estado del tránsito vehicular.

(*) GUDA-Grupo Universitario de Automatización. Universidad Tecnológica Nacional - Facultad Regional Resistencia French 414. 3500 Resistencia-Chaco. República Argentina

Clasificador Bayesiano

El clasificador Bayesiano Normal Naïve (o ingenuo). Es "ingenuo" porque asume que todas las características son independientes (Zhang H., 2004). Es un clasificador eficaz que puede manejar varios estados del sistema. Este clasificador es el caso más simple posible de lo que ahora es una gran y creciente campo conocido como redes bayesianas, o modelos gráficos probabilísticos.

Árbol de Decisiones

Se basa en proporcionar un conjunto de reglas que se construyen internamente con los datos del tráfico (Liu, B, 2007) y así se determina el estado del mismo, a través de la determinación de ciertos nodos y ramas, las cuales representan determinadas condiciones lógicas que se evaluarán en contraste con los datos del tránsito empleado.

Redes Neuronales

Este tipo de clasificador se basa en la asignación de pesos a nodos de una red que se construye en capas, lo que la hace muy precisa, pero consume más recursos del sistema si no se la construye eficientemente (Liu, B, 2007), en nuestro caso solo se necesitó una capa debido a que la simulación no ofrece mucha dificultad para la red.

Visión Artificial

La herramienta de visión artificial elegida se denomina OpenCV (Bradski G., Kaebler A., 2008). Tiene incorporada una biblioteca de procesamiento digital de imagen o PDI. Proporciona un entorno de desarrollo fácil de utilizar y altamente eficiente en tareas de investigación y desarrollo. La librería de OpenCV tiene incorporada fun-

ciones de PDI que facilitan la extracción de información en fotogramas. Dichas rutinas fueron desarrolladas a bajo nivel permitiendo velocidades de procesamiento elevadas.

METODOLOGÍA

El enfoque metodológico utilizado para desarrollar un algoritmo que permita establecer el estado del tránsito vehicular es cualitativo. De esta manera, a partir de los datos extraídos de fotogramas asociados a algoritmos de ML como por ejemplo: Bayes-Naïve, Árbol de decisión y Redes Neuronales permitirán obtener predicciones del estado del tránsito complejo.

PROCEDIMIENTO

Se implementa un software de alto nivel articulado con herramientas de PDI. De esta manera se logra desarrollar un programa encargado de simular el funcionamiento de un sensor del tipo Laser Sicks PLS. En la figura 1 muestra al sensor analizando un cruce de calle.

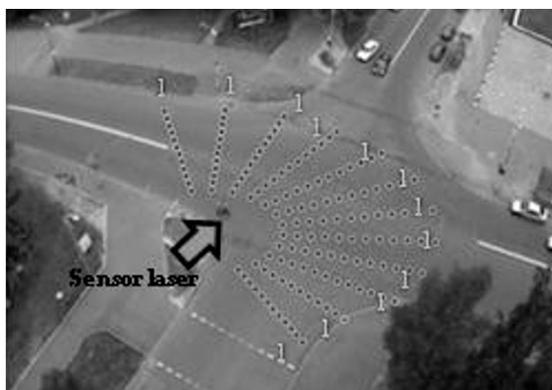


Figura 1: Simulación del sensor laser

Se observan 12 segmentos separados 15°. Cuando un vehículo o peatón interrumpe un segmento, el valor mostrado

cambia de 1 a 0. Si no hay circulación vehicular, la longitud de los doce segmentos es iguales a 1. Es decir, el fotograma de la figura 1, representa un vector de doce columnas iguales a 1. Por ejemplo, en la *figura 2* se visualiza el fotograma 9 donde un peatón

interrumpe el segundo segmento y un vehículo corta los segmentos seis y siete.

Los vectores representativos de un estado determinado del tránsito son almacenados en un archivo de texto como muestra la *Tabla 1*.

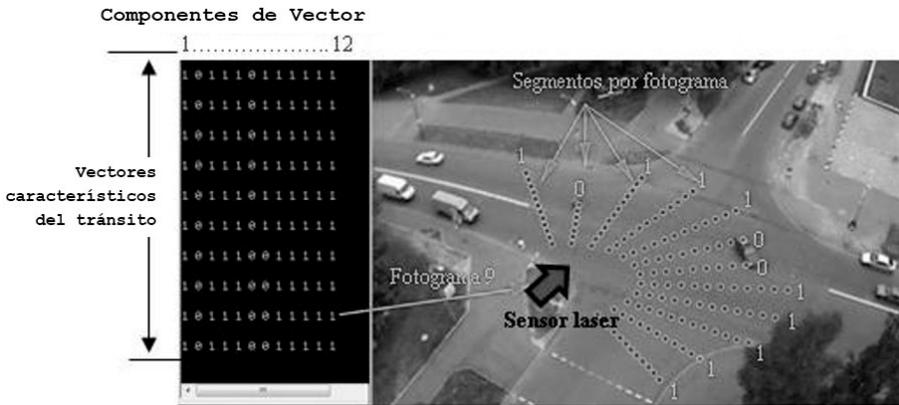


Figura 2: Representación de vectores característico del tránsito

Tabla 1. Representación de vectores característico del tránsito

Estado	Datos del sensor Laser										
A	1	1	1	1	1	1	1	1	1	1	1
B	0	0	1	1	1	1	1	1	1	1	1
B	1	0	0	1	1	1	1	1	1	1	1
B	1	1	0	0	1	1	1	1	1	1	1
B	1	1	1	0	0	1	1	1	1	1	1
B	1	1	1	1	0	0	1	1	1	1	1
B	1	1	1	1	1	0	0	1	1	1	1
B	1	1	1	1	1	1	0	0	1	1	1
B	1	1	1	1	1	1	1	0	0	1	1
B	1	1	1	1	1	1	1	1	0	0	1
B	1	1	1	1	1	1	1	1	1	0	0
C	0	0	0	1	1	1	1	1	1	1	1
C	1	0	0	0	1	1	1	1	1	1	1
C	1	1	0	0	0	1	1	1	1	1	1
C	1	1	1	0	0	0	1	1	1	1	1
C	1	1	1	1	0	0	0	1	1	1	1
C	1	1	1	1	1	0	0	0	1	1	1

Los estados del tránsito vehicular se dividen en cuatro categorías. Se expresan mediante letras y están ubicadas en la primera columna de la *Tabla 1*. La letra **A** representa tránsito nulo, **B** tránsito bajo, **C** tránsito medio y **D** tránsito congestionado.

Se organizan los valores de la *Tabla 1* en una matriz de entrenamiento, definida en el lenguaje de programación utilizado (Python, C, C++, Java). Luego se crean las instancias de archivos de configuraciones de los distintos algoritmos y se establecen los parámetros de los clasificadores de Árboles de decisión, Naïve-Bayes y Redes neuronales.

Se realiza el entrenamiento utilizando la función `train()` para cada uno de los algoritmos de clasificación (Bradski G., Kaebler A., 2008).

Finalmente se obtiene los tres archivos de aprendizaje que permitirán predecir el estado del tránsito complejo.

El esquema representativo de la experiencia se muestra en la *figura 3*.

Los programas desarrollados en este trabajo pueden ser obtenidos en (Vázquez R. D., Fernández J. M., Simón F., Olivetti A. B., 2017) donde además se pueden hacer comentarios y sugerencias.

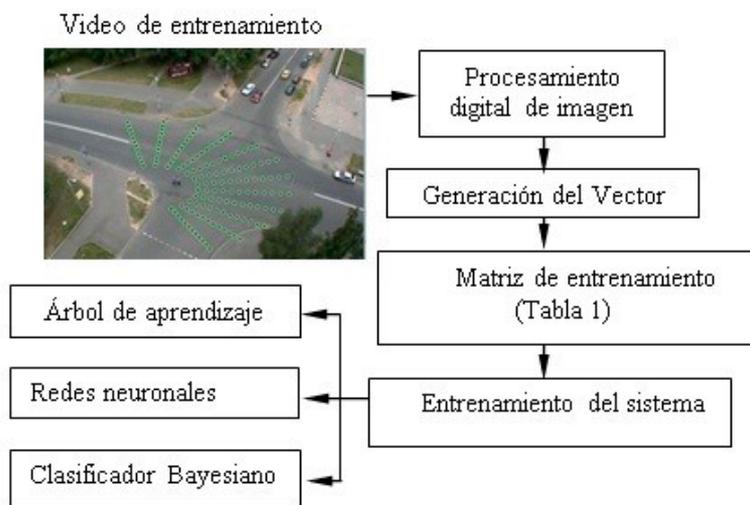


Figura 3: Esquema representativo del entrenamiento y generación del archivo de aprendizaje

DETERMINACION DE LA TASA DE ACIERTO DE LOS ARCHIVOS DE APRENDIZAJE

Con la finalidad de determinar la confiabilidad de los predictores se realizan

pruebas de estimación de la tasa de acierto. Para ello se utilizan un conjunto de vectores que representan diversos estados del tránsito conocidos. Luego se cargan esos vectores al predictor, y se comparan las predicciones como se muestra en la *figura 4*.

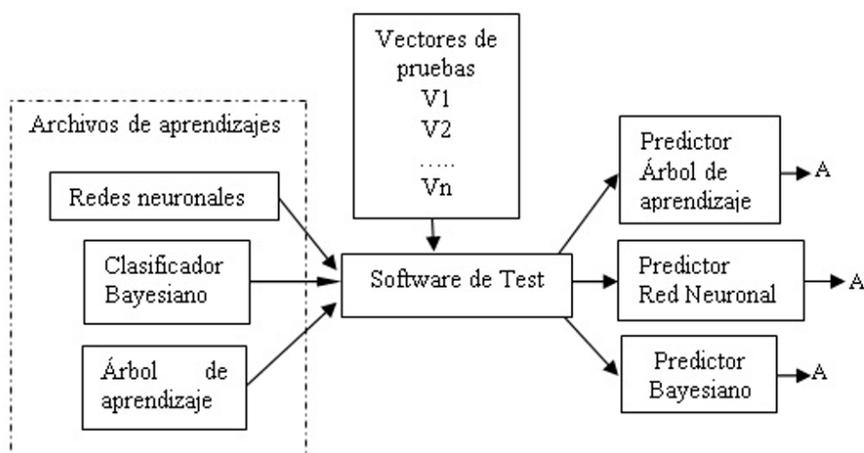


Figura 4: Esquema representativo en la estimación en la tasa de acierto

De esta manera, se obtiene una tasa de acierto para cada uno de los predictores asociados a los modelos predictivos. Permitted realizar comparaciones y seleccionar los más conveniente.

de decisión y la red Neuronal, fueron muy similares. Obteniendo el mejor promedio el árbol de decisión, con un 80,53%. Esto se debe a cómo funciona el modelo de Bayes con respecto a los demás modelos.

RESULTADOS Y DISCUSIÓN

En una primera fase se utilizaron los datos de la *Tabla 2* en el entrenamiento.

Tabla N°2: Número de muestras

Datos Entrenamiento	96
Datos Test (seleccionadas aleatoriamente)	25
Total muestras	121

Una vez seleccionado los datos de entrenamiento y de prueba, se realizaron quince ensayos utilizando el software desarrollado. Los resultados obtenidos se visualizan en la *Tabla 3*. La primera columna representa los test. Los tres restantes muestran el porcentaje de acierto de cada uno de los modelos.

Finalmente se obtiene una tasa promedio de acierto. Los resultados para el árbol

N° Prueba	Baves Naïve	Árbol de decisión	Red Neuronal
1	64%	76%	80%
2	64%	88%	84%
3	60%	84%	76%
4	76%	92%	88%
5	52%	88%	84%
6	88%	76%	80%
7	68%	88%	76%
8	64%	84%	84%
9	64%	60%	68%
10	56%	76%	60%
11	56%	76%	72%
12	80%	92%	88%
13	32%	72%	84%
14	68%	76%	72%
15	68%	80%	84%
Promedio	64%	80,53%	78,67%

Tabla N°3: Porcentaje de acierto de los modelos predictivos y tasa de acierto.

El mismo se lo usa principalmente para analizar grandes volúmenes de datos, mientras que el modelo de Árboles y Redes Neuronales pueden configurarse para trabajar con menores lotes de información.

Las *figuras 5 y 6*, son una representación gráfica de los datos de la *tabla 3*. Se aprecia ciertas oscilaciones en la tasa de acierto en-

tre pruebas. Es debido a la elección de los datos de entrenamiento y de testeo. En el mejor de los casos, fue una tasa de acierto del 92% para el árbol de decisión, sien-

do éste considerado como un número muy bueno para una predicción. Es decir que, de diez predicciones realizadas, solamente una habrá fallado.

Tasa de acierto en diferentes ejecuciones

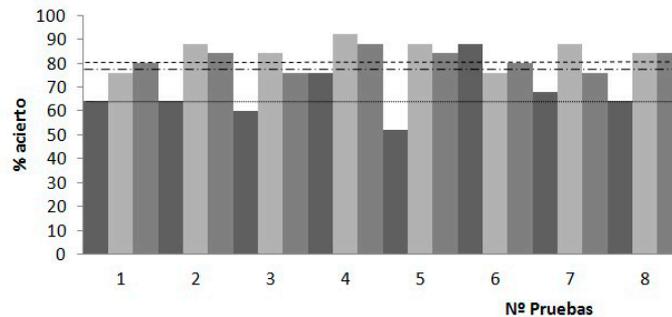


Figura 5: Gráfico representativo de la tasa de aciertos de cada uno de los modelos predictivos

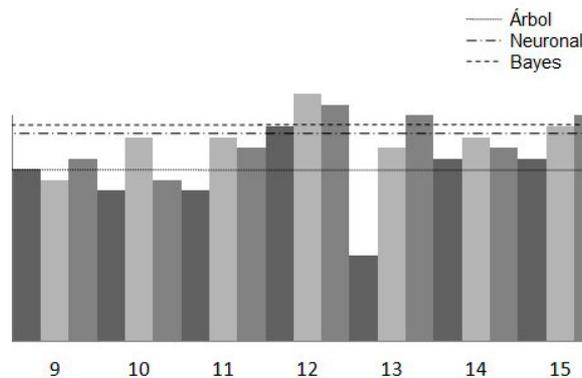


Figura 5: Gráfico representativo de la tasa de aciertos de cada uno de los modelos predictivos

CONCLUSIONES

Fue posible desarrollar un procedimiento para predecir el estado del tránsito empleando herramientas de ML.

Usando dichas herramientas la extracción de información relevante del tránsito vehicular se logró mediante la simulación de un sensor laser y se agruparon los datos representativos del estado vehicular en forma de vectores. Luego se clasificaron

en cuatro categorías ellos fueron: tránsito nulo, tránsito bajo, tránsito medio y tránsito congestionado. En el siguiente paso se usaron estos datos en el entrenamiento de los predictores según las clases. Finalmente se obtuvo 1 archivo de aprendizaje que permitió establecer el estado del tránsito vehicular.

La tasa de acierto del predictor dependió del tamaño de la población y de la habi-

lidad del programador para elegir los ejemplos representativos del sistema vehicular.

En una segunda experiencia se pueden reemplazar los valores obtenidos del simulador por datos reales enviados por sensores que estén ubicados en las esquinas de las calles, debido a los archivos de aprendizajes generados previamente.

Para terminar, se sugiere que la elección del modelo predictivo a usar sea un punto importante de análisis al decidir cómo estructurar la aplicación, siendo que los 3 modelos básicos que propone *opencv* son válidos para aplicar en cualquier aplicación de análisis de datos con ML, se debe tener en cuenta la velocidad y memoria del dis-

positivo, y de cuantos datos se habrán de analizar o predecir por segundo.

ABSTRACT

It develops a procedure to predict the transit status of a vehicular system using as tools: generic devices, simulation of a laser sensor and Machine Learning. This prediction is made through three models: Bayer-Navier Classifier, Decision Tree and Neural Networks. For this it was necessary to perform a training stage prior to the execution of each algorithm. Then a high-level program articulated with the OpenCV library is responsible for merging the data obtained from the vehicular environment and making the prediction.

REFERENCIAS

- Bradski G., Kaehler A., 2008, "Learning OpenCV", Editorial RepKover. Cap 13.pp513. Primera Edición, New York, United States of America.
- Duda R. O., Hart P. E., Stork D. G., 2001, "Pattern Classification", Wiley , New York.
- Duda R. O., Hart P. E., 1973, "Pattern Recognition and Scene Analysis", Wiley, New York.
- Hastie T., Tibshirani R., and Friedman J., 2001, "The Elements of Statistical Learning: Data Mining, Inference and Prediction", Springer-Verlag, New York.
- Liu, B., 2007, "Web data mining : exploring hyperlinks, contents, and usage data", Springer Berlin Heidelberg, New York.
- Vázquez R. D., Fernández J. M., Simón F., Olivetti A. B., 2017, "Predicción del estado del tránsito ML", Universidad Tecnológica Nacional, Facultad Regional Resistencia. <<https://sourceforge.net/projects/prediccion-del-transito/>>
- Zhang H., 2004, "The optimality of naive Bayes," Proceedings of the 17th International FLAIRS Conference.