Congr. Int. en Ing. Electrónica. Mem. ELECTRO, Vol. 43 pp.7 - 10, Oct 2021, Chihuahua, Chih. México
http:// electro.itchihuahua.edu.mx/revista
ISSN 1405-2172

ANÁLISIS DE FACTORES DE DESEMPEÑO EN MÁQUINAS DE VECTOR SOPORTE (SVM) PARA LA DETECCIÓN DE PEATONES

Ruiz Varela Oscar Ramsés, González Rojo Sergio A.
Tecnológico Nacional de México/Instituto Tecnológico de Chihuahua
Tecnológico Nacional de México/Instituto Tecnológico de Chihuahua
Ave. Tecnológico #2909, Chihuahua, Chih., México. C.P. 31310
+51 (614) 2-01-2000
oscar.rv@chihuahua.tecnm.mx, sergio.gr@chihauhua.tecnm.mx

RESUMEN.

El desarrollo de vehículos autónomos como medio para transporte de personas y de mercancías obliga a actualizar las tecnologías de detección de peatones. Las máquinas de vector soporte (SVM) siguen usándose individualmente o en combinación con otras tecnologías (RNN,CNN) para detección de peatones. La versatilidad de las SVM y su posible desempeño de hasta 99% en precisión de reconocimiento, además del bajo requerimiento de poder de cómputo las mantienen en desarrollo. Sin embargo existen factores que inciden directamente en su desempeño en la detección de peatones, tales como el número de muestras de entrenamiento, el tipo de kernel utilizado, la calidad de las muestras tanto en homogeneidad como en su enfoque. Se presenta un análisis de los resultados en el entrenamiento y predicción de SVM en condiciones controladas (datos de entrenamiento
- dataset - seleccionado, número de muestras, tipo de kernel, tiempo de entrenamiento y predicción) usadas en la detección de peatones.

Palabras Clave: SVM, factores de desempeño, detección de peatones.

ABSTRACT.

Development of autonomous vehicles used as person and merchandise way of transport, requires that pedestrian detection technologies keep updated. Support vector machines keeps being used standalone or combined with other technologies (RNN, CNN) for pedestrian detection. SVM versatility and 99% precision achievable, beside its low computing power requirements keeps it on demand. However there are aspects that affects its performance at pedestrians detection, such as samples number for training, kernel type used, samples quality regarding homogenity and orientation. Its presented an analysis of results in SVM training and predictions in controlled conditions (chosen training data –dataset-, number of samples, kernel type, time o training and predictions) used in pedestrian detection.

Keywords: SVM, performance factors, pedestrian detection.

1. INTRODUCCION

Las tecnologías de inteligencia artificial han aprovechado los recursos disponibles en diversas áreas de la ciencia (probabilidad, estadística -muestreo-, algebra lineal -matrices, transformaciones lineales, valores propios-, calculo vectorial –gradiente, producto punto -, termodinámica -entropía-, etc.).

En general la detección de objetos se ha logrado por la investigación en enfoques de clasificación, de rasgos, y de manejo de articulaciones. Los enfoques de clasificación utilizados incluyen varios aceleradores de clasificación, clasificadores SVM, modelos gramáticos y modelos profundos. El estudio de rasgos incluye la revisión de rasgos de parecido Haar, edgelets, shapelets, Histograma de gradiente orientado HOG, bolsa de palabras, histogramas integrales, histogramas de color, descriptores de covarianza, etc. [8]

La detección de peatones como el primer paso y el más fundamental, en muchas tareas del mundo real (análisis de comportamiento humano, reconocimiento de andar, vigilancia de video inteligente, conducción vehicular automática) ha atraído atención masiva en la última década. [1]

También se han combinado desarrollos de inteligencia artificial para explotar las fortalezas combinadas, tal como el uso de redes convolucionales como extractor de rasgos, y las SVM como clasificador. [2][5]

El progreso en la detección de peatones no ha mostrado señales de ralentización en los últimos años, a pesar de los grandes avances en desempeño. Se espera lograr desempeños humanos, y posteriormente, puntuaciones sobre humanas.[6]

Durante la última década se han creado varios benchmark para esta tarea. Esos benchmark han permitido un gran progreso en el área. Aunque no está claro que tan bien se traduce este progreso en desempeño para el mundo real. Se propone que es tiempo de dar énfasis no solo al desempeño intra datasets, sino al desempeño entre datasets. [7]

El objetivo del presente trabajo es analizar el efecto de la selección de parámetros para sintonizar una máquina de vectores soporte en la tarea de identificar peatones en ambientes al exterior. También se analiza el efecto de realizar el entrenamiento con imágenes que incluyan únicamente el objeto de interés, procurando eliminar imágenes externas.

Las SVM, toman los vectores de entrada, las imágenes, mapeadas no linealmente en un espacio vectorial de mayor dimensión, a saber los rasgos –features- obtenidos en este caso por medio de Histogramas de Gradiente Orientados (HOG). En este espacio de rasgos, se construye una superficie lineal de decisiones. Las propiedades especiales de esta superficie de

decisión aseguran una alta habilidad de generalización, propia del aprendizaje de máquina.[2]

Si se considera el híper plano general como indica la ec. 1:

$$w_0 * z + b_0 = 0 \qquad \text{ec. 1}$$

Donde w_0 es el vector de pesos seleccionados, z el vector de rasgos y b_0 el bias o factor de compensación. De la ec. 1 podemos asumir que los pesos son la sumatoria de vectores soporte modificados por α, siendo α un escalar, como indica la ec. 2:

$$w_0 = \sum_{vectores\ soporte} w_i\, z_i \qquad \text{ec. 2}$$

La función de decisión lineal I(z) en el espacio de rasgos será como indica la ec. 3:

$$I(z) = sign(\sum_{vectores\ soporte} \alpha_i\, z_i * z + b_0 \quad \text{ec. 3}$$

El problema conceptual que resuelven las SVM es como encontrar un híper plano separador que generalice bien, fig 1. La dimensionalidad del espacio de rasgos será enorme, y no todos los híper planos que separan los datos de entrenamiento necesariamente generalizan bien. [3]
El problema técnico en los híper planos es como computacionalmente tratar tales espacios altamente dimensionales: para construir polinomios de grado 4 o 5 en un espacio de 200 dimensiones puede ser necesario construir híper planos en un espacio de billones de dimensione.
La parte conceptual de este problema fue resuelto en 1965 para el caso de híper planos óptimos para clases separables. Un híper plano óptimo se define como la función decisión lineal con el máximo margen entre los vectores de las dos clases. Se ha observado que para construir tales hiper planos óptimos se tiene que considerar solo una pequeña cantidad de datos, los llamados vectores de soporte, a saber parte de los datos de entrenamiento son los vectores de soporte.[3]

2. DESARROLLO.

No se definieron las limitaciones para descartar imágenes como peatones en este análisis, como lo sugiere por ejemplo el estándar del dataset Caltech, donde se establece que las personas con oclusiones mayores a 0.5, o de altura menor a 50 pixeles son ignoradas. [4]
Se consideraron los siguientes datasets de peatones: Daimler, INRIA, Fundan y Personal.
El dataset Daimler tiene algunos aspectos que mencionar. Primeramente se usa un formato de imagen con extensión ".pgm", que no es tan popular como ".jpg" o ".png".
Al no encontrar un programa para leer ese formato, se construyó una pequeña aplicación usando la librería PIL del ambiente python, que convierte las imágenes a formato ".png".
Se trabajó con conjuntos de 300 imágenes positivas por dataset, salvo que se indique otra cantidad.
El segundo aspecto a mencionar es que el dataset se compone de imágenes extraídas de un video filmado con una cámara montada en un automóvil, por lo que las imágenes no siempre contienen peatones. Las imágenes que contienen peatones siempre tienen parte del ambiente.
El último aspecto a considerar es que se tienen imágenes claras, sin movimiento en las tomas.
Los datasets INRIA y Fundan se combinaron, para reunir las 300 imágenes. Son dataset típicos. Todas las imágenes contienen peatones en primer plano, de cuerpo completo, en diversas poses. Las imágenes contienen parte del entorno, además del peatón.
El dataset Personal se construyó por medio de imágenes recuperadas de la web. Se escribió un programa para extraer partes de las imágenes, explícitamente del torax hacia arriba, de peatones, ciclistas y motocicilistas. La justificación de usar solo la parte alta del cuerpo fue que en el entrenamiento no se quería incluir el "ambiente" en la captura de los objetivos, para evitar que la SVM genere rasgos que no sean del individuo, y con ello agregar datos que son extraños, para luego depurar esos datos extraños en otro proceso. Zhang y otros estudian el impacto del ruido en el entrenamiento, y concluyen que se obtiene mejoras aun con una pequeña porción de datos de entrenamiento limpios. [6]
Como parte del pre procesamiento se igualaron las imágenes al tamaño de 640x480 pixeles, para permitir la generación del mismo número de rasgos en cada imagen.

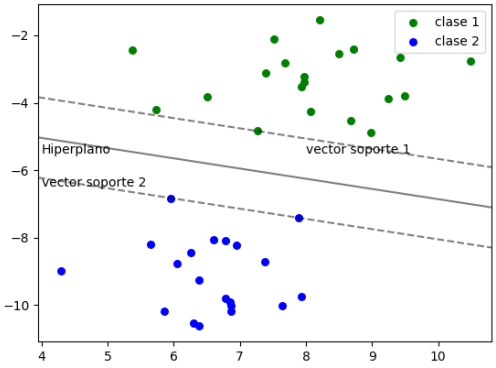

Fig 1. Ejemplo de hiper plano.

Las dos aplicaciones adicionales que se construyeron, una para el entrenamiento y la otra para la prueba, incluían la funcionalidad de registrar aspectos que se consideraron importantes para el análisis posterior, tales como tiempo de duración en la ejecución de la prueba, fechahora, carpetas de origen y destino, resultados de precisión, recall, score F1, valor del parámetro C, tipo de kernel utilizado, numero de muestras positivas y negativas, archivo de entrenamiento. Esto se almacenó en archivos de tipo csv.
Los archivos csv contienen 12y 13 columnas.

La información que se muestra a continuación en tablas y figuras proviene de los archivos csv obtenidos.

3. RESULTADOS DEL ANÁLISIS.

Tabla 1. Comparación tiempo de entrenamiento.

C	NumP	NumN	SVM	tiempo
1	350	268	Lineal	0h 14m 37s
5	350	268	Lineal	0h 17m 2s
10	350	268	Lineal	0h 17m 5s
20	350	268	Lineal	0h 5m 48s
30	350	268	Lineal	0h 13m 16s

La tabla 1 muestra los tiempos de entrenamiento para cinco pruebas usando SVM lineal, con 350 muestras positivas y 268 negativas. La variación del parámetro C nos afecta en los tiempos de entrenamiento desde 5:48 hasta 17:5.
En las SVM, el parámetro C controla el número y la severidad de las violaciones del margen y del híper plano que se toleran en el proceso de ajuste. Si C es infinito no se permite ninguna violación del margen.

Tabla 2. Efecto del parámetro C en tiempos de prueba.

NumPos	NumNeg	FalsosP	FalsosN	Preci	tiempo
288	189	15	11	0.94791667	0h 8m 49s
288	187	15	11	0.94791667	0h 10m 34s
288	187	15	11	0.94791667	0h 2m 29s
288	187	15	11	0.94791667	0h 2m 4s
288	187	15	11	0.94791667	0h 6m 40s

La tabla 2 tiene relación directa con la tabla 1. El renglón 1 de la tabla 1 usó un valor de C=1 para entrenar. El renglón 1 de la tabla 2 muestra que la prueba tomó 8:49. Pero los Falsos positivos y negativos fueron idénticos en las cinco pruebas, misma razón por la que la precisión fue la misma. Sin embargo, con el valor de C=20 de la tabla 1, cuarto renglón, el entrenamiento tuvo el menor tiempo, de 5:48, y con ese mismo valor de C, en la tabla 2, el tiempo de prueba tomó 2:04, que también fue el menor.
El resto de las pruebas se realizaron con C=20 por esa ventaja en los tiempos.

Tabla 3. Entrenamiento con diferentes datasets

ArchivoPkl	NumP	NumN	SVM	tiempo
person20.pk	300	268	linear	0h 14m 23s
person21.pk	301	268	linear	0h 13m 54s
person22.pk	301	268	linear	0h 6m 36s

La tabla 3 muestra la columna ArchivoPkl, que es el archivo que guarda los parámetros del entrenamiento. Con este archivo realizamos un entrenamiento, o dicho de otra manera, encontramos el mejor híper plano, el cual se compone de la colección de pesos que se calcularon, y lo almacenamos. Cuando hacemos la prueba, leemos ese archivo pkl y lo usamos para evaluar las imágenes y obtener un valor de -1 o +1, que nos indica que la imagen contiene o no peatones.
La única observación es mencionar que el menor tiempo se tuvo con el renglón de person22, que es el dataset construido con bustos. Aunque el tamaño es el mismo de todas las imágenes, la imagen es menos variada.

Tabla 4. Prueba en dataset Daimler, diferente entrenamiento

ArchivoPkl	NumPos	NumNeg	FalsosP	FalsosN	Preci	tiempo
person20.pk	500	185	68	47	0.864	0h 11m 30s
person21.pk	500	185	0	11	1	0h 11m 9s
person22.pk	500	185	84	24	0.832	0h 13m 46s

El entrenamiento de person21 con el dataset INRIA obtuvo la mejor precisión, pero llama la atención que tiene falsos negativos, es decir, 11 imágenes donde se detectaron peatones y no los había. En otra vista de la Tabla 4, que llamamos Tabla 4a, se observa que el renglón de person21 tiene un recall de 0.978 en lugar del recall perfecto de 1.

Tabla 4a. Prueba en dataset Daimler, diferente entrenamiento

ArchivoPkl	FalsosP	FalsosN	Preci	Recall	Score
person20.pk	68	47	0.864	0.90187891	0.8825332
person21.pk	0	11	1	0.97847358	0.98911968
person22.pk	84	24	0.832	0.94545455	0.88510638

Tabla 5. Entrenamiento lineal+SDG con diferentes dataset

ArchivoPkl	NumP	NumN	SVM	tiempo
hiper23.pkl	300	268	lineal	0h 2m 25s
hiper24.pkl	301	268	lineal	0h 2m 28s
hiper25.pkl	301	268	lineal	0h 3m 2s

Este entrenamiento usó Gradiente estocástico descendente (SDG), el cual se implementó a bajo nivel, sin usar librerías para ese propósito. Los tiempos de entrenamiento fueron los menores hasta ahora.

Tabla 6. Prueba en diferentes dataset mismo entrenamiento

ArchivoPkl	NumPos	NumNeg	FalsosP	FalsosN	Preci	tiempo
hiper23.pkl	500	185	76	63	0.848	0h 10m 26s
hiper24.pkl	500	185	1	16	0.998	0h 9m 59s
hiper25.pkl	500	185	107	48	0.786	0h 9m 37s

Dado que el entrenamiento se hizo a bajo nivel, la prueba o predicción se realiza únicamente con un producto punto entre la muestra en evaluación y el conjunto de pesos producto del entrenamiento, por lo que la prueba es un poco más corta, comparando con los tiempos de la tabla 4.

El mejor resultado, del archivo hiper24, es el entrenamiento con el dataset INRIA y la prueba con el dataset Daimler.

Tabla 7. Entrenamiento con diferentes datasets. Igual que tabla3.

ArchivoPkl	NumP	NumN	SVM	tiempo
person26.pk	300	268	linear	0h 13m 33s
person27.pk	301	268	linear	0h 11m 40s
person28.pk	301	268	linear	0h 5m 26s

La variación en los tiempos de las tablas 3 y 7 se explica por la ejecución de programas adicionales durante el entrenamiento.

Tabla 8. Prueba en dataset personal, diferente entrenamiento

ArchivoPkl	NumPos	NumNeg	FalsosP	FalsosN	Preci	tiempo
person26.pk	300	185	51	47	0.83	0h 2m 17s
person27.pk	300	185	15	11	0.95	0h 2m 10s
person28.pk	300	185	2	24	0.9933	0h 2m 42s

Tiempos muy cortos en la evaluación del dataset personal.
El mejor puntaje se obtiene para el dataset INRIA.

4. CONCLUSIONES

El grado de complejidad de las imágenes en el dataset modifica el resultado de la evaluación.
La configuración de los parámetros en ambas operaciones, entrenamiento y prueba, pueden alterar los tiempos sin entregar necesariamente mejores resultados.
Existe un grado de incertidumbre respecto a que resultados se obtendrán dependiendo de las características dominantes en las imágenes del entrenamiento respecto a las imágenes de prueba.

El ejemplo que se utilizó, la detección de peatones para realizar este análisis es sencillo, al menos sin considerar las restricciones de una detección en tiempo real, donde la velocidad del vehículo, la velocidad del peatón, los puntos ciegos, las oclusiones, demandan más que la identificación correcta.
Otro factor a considerar es distinguir una persona de un maniquí, una fotografía de cuerpo completo, u otro seudo peatón.

Complementario a este análisis, se puede realizar una evaluación del dataset previamente,

Referencias.

[1] J.Mao, T. Xiao, Y. Jiang, Z. Cao. What can help pedestrian detection?. CVPR 2017.
[2] N. Xiao Xiao, S.Y. Ching. A novel hybrid CNN-SVM classifier for recognizing handwritten digits. Pattern recognition. Vol 45,Issue 4, Abril 2012. Pag 1318-1325.
[3] C. Cortes, V. Vapnik. Support vector networks. Machine learning. Vol 20, 1995. Pag 274,276
[4] P. Dollar, C. Wojek, B. Schiele, P. Perona . A benchmark in computer vision and pattern recognition . CVPR .2009. pages 304-311
[5] W. Ouyang, X. Wang. , Joint deep learning for pedestrian detection, ICCV 2013.pages 2056-2063.
[6] S. Zhang, R.Benenson, M. Omram, J. Hosang,B. Schiele. How far are we from solving pedestrian detection?. CVPR 2016. Pages 1259-1267
[7] S. Zhang, R.Benenson, M. Omram, J. Hosang,B. Schiele. CityPersons: A diverse dataset for pedestrian detection. Proceedings of the IEEE Conference on computer vision and pattern recognition (CVPR). 2017. Pages 3213-3221
[8] W. Ouyang, X. Wang et al. Single pedestrian detection aided by multi pedestrian detection. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2013. Pages 3198-3205