

Abstract

The recognition of images through deep learning, has taken force in recent times, especially when the evolution in the computational mechanisms has increased, this advantage makes that execute algoritmos of learning is faster to execute and its efficiency is not seen reduced. we thus see the use of these techniques used either for the recognition of images, as well as processing transactions of doubtful order at the banking level. The present article deals with the use of a neural network to analyze images of numbers written by hand, with the use of an API of the Google company known as TensorFlow.

INTRODUCCIÓN

El Presente artículo está enfocado al reconocimiento de imágenes por medio del aprendizaje profundo, el cual busca por medio de la construcción de grafos, generar condiciones de aprendizaje que permite determinar por medio de modelos suministrados a la red, el reconocimiento voz, representaciones gráficas, detención de objetos entre otros dominios.

Empezaremos hablando de algunos conceptos relacionados con el Deep learning como se le conoce en el idioma inglés, el planteamiento de lo que queremos lograr, mostraremos unos resultados y las conclusiones finales.

PLANTEMIANTO DEL PROBLEMA

Uno de los grandes retos en el día de hoy es reconocer estructuras o patrones en diversas situaciones, como reconocer un rostro sin importar el ángulo, contar objetos presentes en una imagen para

inventarios, controlar tráfico entre otras aplicaciones. Para el presente ejercicio se ha utilizado un grupo de imágenes de números del 0 al 9 escritos a mano, los cuales se usarán como entrenamiento de la red neuronal, para luego con un conjunto de datos de prueba, permitir encontrar las coincidencias. Se han utilizado un paquete de datos cargados de la internet que poseen 55.000 imágenes de entrenamiento, 10000 de pruebas y 5000 de validación.

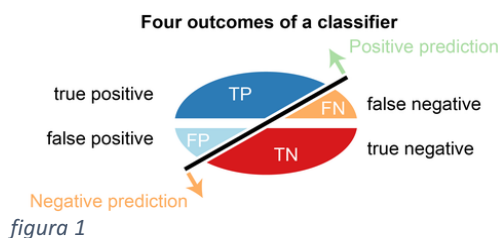
CONCEPTOS

APRENDIZAJE PROFUNDO

El aprendizaje profundo permite modelos computacionales que están compuestos de múltiples capas de procesamiento para aprender representaciones de datos con múltiples niveles de abstracción[1]. Estos métodos han mejorado drásticamente el estado de la técnica en reconocimiento de voz, reconocimiento de objetos visuales, detección de objetos y muchos otros dominios, como el descubrimiento de fármacos y la genómica. El aprendizaje profundo descubre una estructura compleja en grandes conjuntos de datos mediante el uso del algoritmo de retropropagación para indicar cómo una máquina debe cambiar sus parámetros internos que se utilizan para calcular la representación en cada capa a partir de la representación en la capa anterior. Las redes convolucionales profundas han producido adelantos en el procesamiento de imágenes, video, voz y audio, mientras que las redes recurrentes han iluminado los datos secuenciales, como el texto precisa.

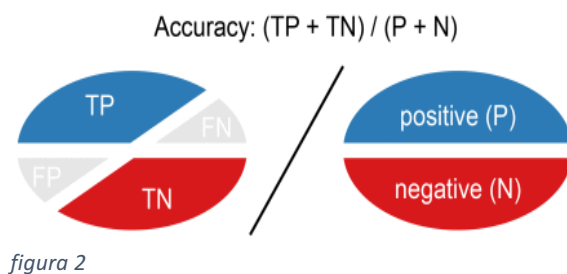
MATRIZ DE CONFUSIÓN

Una matriz de confusión de clasificación binaria es una tabla de dos por dos formada por el recuento del número de los cuatro resultados de un clasificador binario. Normalmente los denominamos TP, FP, TN y FN en lugar de "la cantidad de verdaderos positivos"[2], y así sucesivamente. Como se observa en la figura 1.



PRECISIÓN

La precisión (ACC) se calcula como el número de todas las predicciones correctas divididas por el número total del conjunto de datos. La mejor precisión es 1.0, mientras que la peor es 0.0. También se puede calcular por $1 - \text{ERR}$. Como se muestra en la figura 2



ERROR DE ESTIMACIÓN

La tasa de error (ERR) se calcula como el número de todas las predicciones incorrectas dividido por el número total

del conjunto de datos. La mejor tasa de error es 0.0, mientras que la peor es 1.0.

TENSORFLOW

TensorFlow es una biblioteca de código abierto para aprendizaje automático a través de un rango de tareas, y desarrollado por Google para satisfacer sus necesidades de sistemas capaces de construir y entrenar redes neuronales para detectar y descifrar patrones y correlaciones, análogos al aprendizaje y razonamiento usados por los humanos, esta herramienta es la que ha sido usada en el presente trabajo para construir la red neuronal con la cual se pretende predecir, dada una imagen de un número sea capaz de decir a cual pertenece.

ESTADO DEL ARTE

1. APRENDIZAJE AUTOMÁTICO EN CATEGORIZACIÓN AUTOMATIZADA DE TEXTO

Este artículo aborda La categorización automática de los textos en categorías predefinidas lo cual genera un interés creciente en los últimos años, debido a la mayor disponibilidad de documentos en formato digital y la consiguiente necesidad de organizarlos. En la comunidad investigadora, el enfoque dominante de este problema se basa en las técnicas de aprendizaje automático: un proceso inductivo general construye automáticamente un clasificador al aprender, a partir de un conjunto de documentos preclasificados, las características

de las categorías. Las ventajas de este enfoque sobre el enfoque de ingeniería del conocimiento (que consiste en la definición manual de un clasificador por expertos de dominio) son una muy buena eficacia, ahorros considerables en términos de mano de obra experta y portabilidad directa en diferentes dominios. Esta encuesta discute los principales enfoques de categorización de texto que caen dentro del paradigma de aprendizaje automático. Discutiremos en detalle cuestiones relacionadas con tres problemas diferentes, a saber, la representación de documentos, la construcción de clasificadores y la evaluación de clasificadores[3].

2. REDES CONVOLUCIONALES MUY PROFUNDAS PARA EL RECONOCIMIENTO DE IMÁGENES A GRAN ESCALA

En este trabajo se presenta el efecto de la profundidad de red convolucional en su precisión en el ajuste de reconocimiento de imágenes a gran escala. la principal contribución es una evaluación exhaustiva de las redes de profundidad creciente utilizando una arquitectura con filtros de convolución muy pequeños (3×3), lo que demuestra que se puede lograr una mejora significativa en las configuraciones de la técnica anterior al empujar la profundidad a 16-19 capas. investigaciones sobre el uso de representaciones

visuales profundas en la visión por computadora[4].

PROCEDIMIENTO.

Se tomaron el conjunto de imágenes prueba y de entrenamiento y se cargaron al sistema, si estas no estaban en la ruta destino, la instrucción de tensorflow descargaba desde un repositorio todo lo necesario para entrenar la red.

El conjunto de datos se cargó con la codificación denominada One-Hot. Esto significa que las etiquetas se han convertido de un solo número a un vector cuya longitud es igual a la cantidad de clases posibles. Todos los elementos del vector son cero excepto el elemento i ésimo que toma el valor uno; y significa que la clase es i .

Se creó una función que es utilizada para trazar 9 imágenes en una cuadrícula de 3×3 y escribir las clases verdaderas y predichas debajo de cada imagen, con el fin de comprobar que todo estaba bien.

Se pasó a definir las variables de marcador de posición para las imágenes de entrada (Etiquetas variables), a la que llamaremos x . Hacer esto permite cambiar las imágenes que se ingresan al grafo de TensorFlow.

El tipo de dato que se introducen en el grafo, son vectores o matrices multidimensionales (denotados tensores). Estos tensores son arrays multidimensionales, cuya forma es `[None, img_size_flat]`, donde **None** significa que el tensor puede contener un número arbitrario de imágenes, siendo cada imagen un vector de longitud `img_size_flat`.

Finalmente, se tiene la variable de marcador de posición para la clase verdadera de cada imagen en la variable de marcador de posición x . Estos son enteros y la dimensionalidad de esta variable de marcador de posición se predefine como **[None]**, lo que significa que la variable marcador de posición es un vector unidimensional de longitud arbitraria.

Como se ha indicado anteriormente, en éste ejemplo vamos a emplear un simple modelo matemático de regresión lineal, es decir, que vamos a definir una función lineal donde se multiplica las imágenes en la variable de marcador de posición x por una variable w que llamaremos pesos y luego agrega un sesgo (sesgo) que llamaremos b .

Por tanto:

$$\text{logist} = wx + b$$

Posteriormente se realiza la función de coste y optimización otorgada por el tensor para obtener los resultados mostrados abajo.

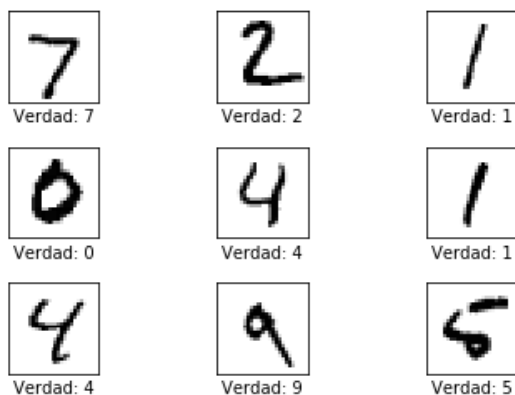


Figura 3, Set de entrenamiento

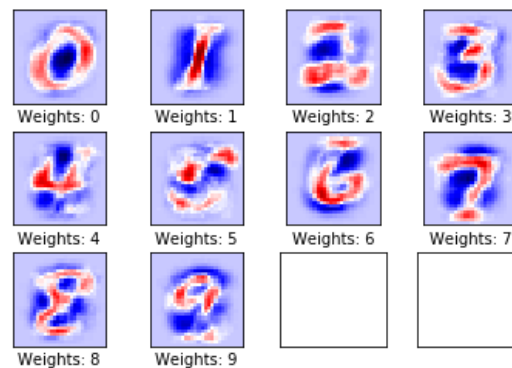


Figura 4 Imágenes predichas

CONCLUSIONES

El modelo de regresión lineal simple que se ha establecido para el presente ejercicio logra en un porcentaje superior al 92% predecir una imagen dada, después de su entrenamiento, empero este solo se logró una vez ejecutada 5000 iteraciones; es un porcentaje bueno, empero se debe mejorar el modelo de tal manera que con menor uso de CPU o GPU, se logró mayor porcentaje de aciertos y menos consumo computacional, dado que a mayor complejidad de lo que se desea predecir mayor es la capacidad de recursos computacionales que se requieren.

REFERENCIAS

- 1.[1]LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
- 2.[2]Gonzalez.JC.. TensorFlow para principiantes 2017: <https://www.apsl.net/blog/2017/12/05/tensor-flow-para-principiantes-i>
- 3.[3]Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.
- 4.[4] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

