

Taller 2
Apoyo a supermercado inteligente
Ciencia de Datos Aplicada

Presentado por:
Adriana María Ríos
Diego Alberto Rodríguez Cruz

Docente(s):
Fabian Camilo Peña
Juan Pablo Reyes
Nancy Marcela Cubillos

Universidad de los Andes
Facultad de Ingeniería
Departamento de Ingeniería de Sistemas y Computación
Maestría en Ingeniería de Información - MINE
2024-II

Entendimiento y preparación de los datos.

Los datos de entrada están compuestos por 5.418 imágenes ubicadas en directorios según la categoría, clase y subclase del producto y las etiquetas de clase y subclase de cada imagen distribuidas en 3 archivos: test.txt, train.txt y val.txt. Se identifican 3 categorías principales de productos: Frutas, Empaques y Vegetales.

Las imágenes son compuestas por el ancho, largo y canal. El canal puede tener valores de 1 para una imagen monocromática y 3 para una imagen a color. Para facilitar el entrenamiento del modelo, es necesario transformar las imágenes en un arreglo simple (unidimensional) para lo cual se utilizará la siguiente estrategia:

1. Reducir y estandarizar el tamaño de las imágenes: se utiliza la función *cv2.resize* de *OpenCV* que permite hacer escalamiento de las imágenes, en este caso se utiliza para reducir el tamaño de cada imagen a 128 x 128 píxeles.
2. Aplanar la imagen: se utiliza la función *reshape* de arreglos numpy para convertir la imagen en un arreglo simple calculado a partir de (ancho x largo x canal)
3. Normalizar la imagen: Se dividen todos los píxeles del arreglo entre 255 de manera que el valor de cada píxel variará entre 0 y 1.

Dimensiones del dataset:

Con el fin de identificar las dimensiones de los datos, las características de las imágenes y los indicadores se unificaron en un solo dataset las etiquetas y las imágenes (cargadas previamente en memoria).

El dataset unificado (etiquetas + imágenes) contiene 5.418 registros y 8 columnas, cada registro contiene información de la imagen como ubicación (filename), clase (class_id), subclase (coarse_class_id), tipo de registro train, test o val (mode), tamaño de la imagen en x (size_x), tamaño de la imagen en y (size_y), número de canales (channels) y tamaño total de la imagen (size).

| | filename | class_id | coarse_class_id | mode | size_x | size_y | size | channels |
|---|---|----------|-----------------|-------|--------|--------|--------|----------|
| 0 | train/Fruit/Apple/Golden-Delicious/Golden-Deli... | 0 | 0 | train | 348 | 348 | 363312 | 3 |
| 1 | train/Fruit/Apple/Golden-Delicious/Golden-Deli... | 0 | 0 | train | 348 | 348 | 363312 | 3 |
| 2 | train/Fruit/Apple/Golden-Delicious/Golden-Deli... | 0 | 0 | train | 348 | 348 | 363312 | 3 |

El dataset distribuye el total de los datos en entrenamiento, prueba y validación de la siguiente manera:

| Modo | Cantidad | Porcentaje |
|--------------|--------------|-------------|
| train | 2.639 | 49% |
| test | 2.484 | 46% |
| val | 295 | 5% |
| Total | 5.418 | 100% |

Aproximadamente el 95% de los datos se distribuye en partes iguales para entrenamiento y prueba. El 5% de los datos se reservan para la validación.

Características de las imágenes

El dataset lo compone 5.418 imágenes en total, contiene imágenes de diferentes tamaños, en promedio el tamaño de las imágenes es de 348 x 357 pixeles, sin embargo, este valor puede variar entre 348 x 348 pixeles y 464 x 464 pixeles. Las imágenes ocupan un tamaño promedio en disco de 373.661 bytes. Todas las imágenes son a color ya que el promedio de canales (channels) es 3 y su desviación estándar es 0.

| | size | size_x | size_y | channels |
|-------|---------------|-------------|-------------|----------|
| count | 5418.000000 | 5418.000000 | 5418.000000 | 5418.0 |
| mean | 373661.049834 | 348.792174 | 357.120709 | 3.0 |
| std | 33858.825929 | 9.554136 | 31.224907 | 0.0 |
| min | 363312.000000 | 348.000000 | 348.000000 | 3.0 |
| 25% | 363312.000000 | 348.000000 | 348.000000 | 3.0 |
| 50% | 363312.000000 | 348.000000 | 348.000000 | 3.0 |
| 75% | 363312.000000 | 348.000000 | 348.000000 | 3.0 |
| max | 484416.000000 | 464.000000 | 464.000000 | 3.0 |

Análisis de resultados del modelo.

Al probar dos modelos de Machine Learning se obtuvieron los siguientes resultados:

En el modelo de Random Forest

- Una precisión de 0.1162 significa que solo el 11.62% de los productos que el modelo ha clasificado como correctos son realmente correctos.
- Un recall de 0.2596 significa que solo el 25.96% de los productos correctos están siendo identificados correctamente por el modelo.
- Un F1-score bajo como el 0.1527 sugiere que el modelo tiene un desempeño deficiente tanto en precisión como en recall. Es probable que no esté identificando correctamente los productos y que lo haga de manera poco confiable.
- Un ROC AUC de 0.8262 es relativamente bueno, lo que sugiere que el modelo tiene una buena capacidad para distinguir entre las clases en general.

En el modelo de red neuronal

- Un valor de 0.5978 (aproximadamente 60%) es considerablemente más alto que la precisión obtenida con RandomForestClassifier, lo que sugiere que el modelo CNN está realizando una mejor tarea de clasificación. Esta mejora es significativa. Con una precisión de 60%, el modelo sería capaz de identificar correctamente los productos más de la mitad de las veces, lo que puede ser suficientemente bueno para un sistema de supermercado inteligente, aunque aún se pueden mejorar los resultados.

Insights

Las redes neuronales convolucionales (CNN) son especialmente buenas para tareas de visión por computadora, como la identificación de productos en imágenes. La arquitectura CNN es capaz de aprender patrones complejos y características espaciales de las imágenes, lo que probablemente contribuyó a la mejora en precisión (0.5978).

Es probable que el modelo RandomForestClassifier no esté aprovechando correctamente las características visuales de las imágenes, lo que puede deberse a una falta de calidad en los datos (por ejemplo, imágenes borrosas o desequilibrio entre clases/categorías).

Aunque la precisión de la CNN es relativamente alta, aún se puede mejorar ajustando hiper parámetros como el número de capas, el tamaño de los filtros, el aprendizaje y la tasa de regularización.