



Clasificación de frijoles mediante redes neuronales

Neil Otniel Moreno Rivera

Universidad de Guanajuato, no.morenorivera@ugto.mx

Resumen— La investigación se enfocó en la aplicación de tecnología de visión por computadora para distinguir y clasificar diferentes variedades de frijoles secos con características similares, utilizando un conjunto considerable de datos de imágenes y extrayendo una variedad de características para el análisis. Este enfoque tiene el potencial de contribuir al desarrollo de métodos más eficientes en la clasificación y selección de semillas de frijol.

Abstract— The research focused on the application of computer vision technology to distinguish and classify different varieties of dry beans with similar characteristics, using a considerable set of image data and extracting a variety of characteristics for analysis. This approach has the potential to contribute to the development of more efficient methods in the classification and selection of bean seeds.

I. INTRODUCCIÓN

Este texto describe la metodología empleada en una investigación científica centrada en el análisis de siete tipos distintos de frijoles secos. Se seleccionaron estos tipos considerando diversas características como forma, tipo y estructura, tomando en cuenta las condiciones del mercado. Con el objetivo de lograr una clasificación uniforme de las semillas, se implementó un sistema de visión por computadora.

En el proceso de investigación, se capturaron imágenes de 13,611 granos pertenecientes a las siete variedades diferentes de frijoles secos mediante una cámara de alta resolución. Estas imágenes fueron utilizadas para desarrollar un modelo de clasificación. El sistema de visión por computadora aplicó etapas de segmentación y extracción de características a las imágenes de los frijoles. En total, se identificaron 16 características, de las cuales 12 estaban relacionadas con dimensiones y 4 con formas de los granos.

II. TEORIA

El código inicia con la carga de un conjunto de datos desde el archivo CSV denominado 'Bean_Dataset.csv'. Este archivo contiene información detallada sobre siete variedades distintas de frijoles secos, cada una caracterizada por 16 atributos diversos como forma, tipo y estructura. Estos atributos se han seleccionado cuidadosamente por su relevancia en la clasificación de las semillas.

Posteriormente, se lleva a cabo la estandarización de las características de entrada mediante la implementación de la clase `StandardScaler` de `scikit-learn`. Este paso resulta

fundamental para normalizar las variables y garantizar que todas ellas contribuyan de manera equitativa al proceso de entrenamiento del modelo, evitando que características con magnitudes mayores influyan de manera desproporcionada en el aprendizaje.

A continuación, se procede con la codificación de la variable objetivo, que representa la categoría a la que pertenece cada tipo de frijol. Esta codificación numérica se logra mediante el uso de `LabelEncoder` y se complementa con la transformación a un formato de "one-hot encoding" mediante la función `to_categorical` de `Keras`. Este paso es esencial para preparar la variable objetivo de manera adecuada para el entrenamiento de un modelo de clasificación.

Para evaluar de manera efectiva el rendimiento del modelo, el conjunto de datos se divide en conjuntos de entrenamiento y prueba. El 80% de los datos se destina al entrenamiento del modelo, mientras que el 20% se reserva para evaluar su capacidad de generalización.

La construcción del modelo de red neuronal se lleva a cabo utilizando la librería `Keras`, definiendo un modelo secuencial compuesto por tres capas densas. La última capa, con siete unidades y función de activación 'softmax', se adapta a la naturaleza multiclase del problema de clasificación de frijoles secos.

El modelo se compila utilizando el optimizador 'adam' y la función de pérdida 'categorical_crossentropy', configuraciones estándar para problemas de clasificación multiclase. Este paso permite ajustar eficientemente los pesos de la red durante el entrenamiento.

El entrenamiento del modelo se realiza utilizando el conjunto de entrenamiento, optimizando los pesos de la red para minimizar la función de pérdida y mejorar su capacidad de clasificación.

La evaluación del modelo se efectúa en el conjunto de prueba, calculando y mostrando la pérdida y la precisión del modelo en la clasificación de frijoles secos.

Finalmente, se genera y muestra una matriz de confusión detallada que evalúa la capacidad del modelo para clasificar cada tipo de frijol. Esta matriz proporciona información crucial sobre los aciertos y errores de clasificación en cada categoría, siendo fundamental para comprender la robustez del modelo.

* Neil Otniel Moreno Rivera.

III. RESULTADOS

Los resultados obtenidos fueron los siguientes, la precisión se consiguió un 93% al menos en las pruebas automatizadas, esto nos dice que podrá identificar la mayoría de veces los frijoles de manera correcta, también se muestra la matriz de confusión, probando las 7 categorías posibles en este caso son los 7 tipos de frijoles, mostrando un error mínimo en algunos tipos, en cambio en otros no tiene error alguno.

```
Test loss: 0.1924062967300415
Test accuracy: 0.9360998868942261
```

Fig. 1. Imagen demostrativa de la precisión del modelo.

```
Confusion matrix:
[[229  0 17  0  1  2  6]
 [ 0 92  0  0  0  0  0]
 [ 5  0 336  0  6  1  2]
 [ 1  0  0 677  1  9 47]
 [ 1  0  3  1 378  0  7]
 [ 1  0  0  1  0 362 12]
 [ 0  0  2 40  7  1 475]]
```

Fig. 2 Matriz de confusión obtenida.

IV. CONCLUSIONES

Para concluir con esto, puedo decir que se encontró una solución, satisfactoria, aunque no óptima, puesto que se puede seguir perfeccionando e incluso implementarlo de manera práctica para visión por computadora, estandarizando los valores y calibrando lo necesario.