Amazon ML Challenge

Team: Optimus_Prime

Members: Akshay Perison Davis J, Arulpathi A, Mithran M, Sharan S

Problem Statement:

The problem given to us involves creating a machine learning model that can extract important information from product images, such as weight, volume, dimensions, and other physical attributes. This is crucial for sectors like healthcare, e-commerce, and content moderation, where product details are often missing or incomplete. By analyzing images, the model should predict these details in a structured format (e.g., "34 gram", "12.5 cm").

Machine Learning Approach for Extracting Physical Attributes from Product Images:

1. Problem Overview

The task is to extract important physical attributes such as weight, volume, and dimensions from product images. This is essential for industries like e-commerce, healthcare, and logistics, where these details may not always be readily available or structured. The challenge is to develop a solution that processes product images, extracts relevant textual information using OCR (Optical Character Recognition), and identifies the target attributes using Named Entity Recognition (NER).

2. Machine Learning Approach

Step 1: Image Downloading

We start by downloading the product images using the URLs provided in the dataset. A placeholder image is created if the image cannot be downloaded after several retries. Multiprocessing is used to speed up the downloading process.

Step 2: Text Extraction with EasyOCR

Once the images are downloaded, the next step is extracting the text from the images using EasyOCR. This OCR tool reads the text from images and returns the recognized characters.

Step 3: Named Entity Recognition (NER)

The extracted text is then used to train an NER model that can detect and classify different entities (such as weight, dimension, volume, etc.). We use the Spacy library to train this model.

Step 4: Test Data Prediction

Once the NER model is trained, it is used to predict the entities in the test images. The extracted text from the test images is processed by the NER model to find relevant entities. The final predictions are saved in a CSV file, which is later validated using a provided sanity check script.

3. Machine Learning Models Used

EasyOCR:

- Purpose: Extract text from product images.
- Why: Lightweight, easy to use, and supports multiple languages with good accuracy.

Spacy NER:

- Purpose: Identify specific entities like weight, volume, and dimensions in text.
- Why: Spacy is a popular NLP tool with customizable NER pipelines.

4. Experiments Conducted

Experiment 1: Basic OCR Text Extraction

- Objective: Extract text using EasyOCR and clean it.
- Results: Successful extraction with corrections for common OCR errors.

Experiment 2: NER Model Training

- Objective: Train Spacy NER to recognize entities (weight, volume, etc.).
- Results: Trained with 80/20 split over 10 epochs, showing progressive loss reduction.

Experiment 3: Test Data Prediction

- Objective: Evaluate the NER model on unseen data.
- Results: Successfully predicted entities and saved results for evaluation.

5. Conclusion

The proposed approach efficiently extracts relevant physical attributes from product images by combining EasyOCR for text recognition and Spacy NER for entity identification. The model performed well in identifying key entities like weight and volume from the images. However, further improvements could be made by incorporating more advanced error correction in the OCR step and experimenting with other OCR engines for higher accuracy. Additionally, tuning the NER model with more extensive and varied training data could further boost its accuracy in real-world scenarios.

This approach demonstrates the practical application of combining image processing, OCR, and NLP techniques to solve the problem of extracting structured data from product images.