Zero shot Learning: Clasificación de imágenes

Guadalupe Quispe,William Frank
UNI
Lima, Perú
wguadalupeq@uni.pe

Chavez Bruno,Victor Manuel
UNI
Lima, Perú
vchavezb@uni.pe

Stefano Olivieri Romero
UNI
Lima, Perú
solivierir@uni.pe

***Resumen*—**

I. Introducción

Los modelos basados en aprendizaje profundo han alcanzado un rendimiento muy avanzado para tareas de reconocimiento de imágenes y detección de objetos recientemente. Muchos de estos modelos pueden lograr un rendimiento a nivel humano en conjuntos de datos de clasificación de imágenes complejos como ImageNet, que incluye miles de clases de objetos de imagenes diferentes. Sin embargo, estos modelos se basan en el paradigma de entrenamiento supervisado y su rendimiento depende en gran medida de la cantidad de datos de entrenamiento etiquetados. Además, las clases que los modelos pueden reconocer se limitan a aquellas en las que fueron entrenados.

Esto hace que estos modelos sean menos útiles en escenarios realistas donde puede que no haya suficientes imágenes etiquetadas para todas las clases durante el entrenamiento.

Dado que prácticamente no es posible entrenar en imágenes de todos los objetos posibles, queremos que nuestro modelo reconozca imágenes de clases que no vio durante la fase de entrenamiento, aquí es donde se introduce el uso del paradigma de "Zero-shot Learning".

Para el desarrollo del proyecto utilizamos el lenguaje de programación Python en cual usamos las siguientes librerías que serán detallados a continuación:

- Scikit-learn : Es una biblioteca para aprendizaje automático de software libre que incluye varios algoritmos de clasificación, regresión y análisis de grupos
- PyTorch : Es una biblioteca de aprendizaje automático de código abierto basada en la biblioteca de Torch, utilizado para aplicaciones que implementan cosas como visión artificial y procesamiento de lenguajes naturales.
- TensorFlow : Es una librería que puede utilizarse para crear modelos de Deep Learning directamente o utilizando librerías de envolturas que simplifican el proceso construido sobre TensorFlow.

II. Estado del Arte

II-A. An embarrassingly simple approach to zero-shot learning ,Bernardino Romera-Paredes,Philip H. S. Torr

[5]

La clasificación automática es el primer problema considerado en el aprendizaje automático, por lo que se ha estudiado y analizado, dando lugar a una amplia variedad de enfoques de clasificación que han demostrado su utilidad en muchas áreas como la visión computacional y la clasificación de documentos.

Sin embargo, estos enfoques generalmente no pueden abordar escenarios desafiantes en los que pueden aparecer nuevas clases en la etapa de aprendizaje.

II-B. A Review of Generalized Zero-Shot Learning Methods

[6] En este trabajo con los avances recientes en el procesamiento de imágenes y la visión computacional, los modelos de aprendizaje profundo (Deep Learning) han alcanzado una gran popularidad debido a su capacidad para proporcionar una solución integral desde la extracción de características hasta la clasificación. A pesar de su éxito, los modelos Deep Learning tradicionales requieren entrenamiento en una gran cantidad de datos etiquetados para cada clase, junto con una gran cantidad de muestras. En este sentido, es un desafío recolectar muestras etiquetadas a gran escala. Como ejemplo, ImageNet, que es un gran conjunto de datos, contiene 14 millones de imágenes con 21,814 clases en las que muchas clases contienen solo unas pocas imágenes. Además, los modelos Deep Learning estándar solo pueden reconocer muestras pertenecientes a las clases que se han visto durante la fase de entrenamiento y no pueden manejar muestras de clases no antes vistas.

Si bien en muchos escenarios del mundo real, es posible que no haya una cantidad significativa de muestras etiquetadas para todas las clases. Por un lado, la anotación detallada de una gran cantidad de muestras es laboriosa y requiere un conocimiento experto del dominio. Por otro lado, muchas categorías carecen de suficiente data en este escenario nos encontramos en muchas situaciones. Esto sucede cuando se trata de un conjunto creciente de clases, como la detección de nuevas especies de animales,Muestras etiquetadas, por ejemplo, aves en peligro de extinción, u observadas en

progreso, por ejemplo, COVID-19, o no cubiertas durante el entrenamiento pero aparecen en la fase de prueba.

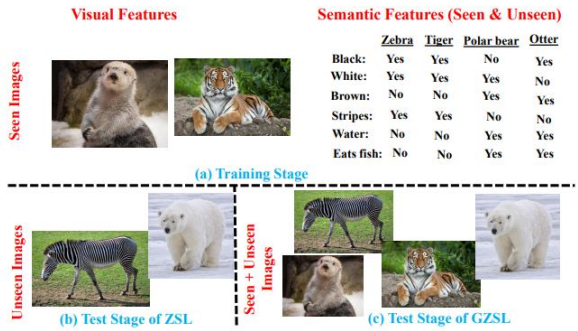

Figura 1. Zero-Shot

II-C. Train Once, Test Anywhere: Zero-Shot Learning For Text Classification

[7]

En este trabajo se propusieron muchos enfoques de zero-shot Learning en el dominio de la visión computacional por Sandouk Chen (2016), Socher et al. (2013). Sin embargo, existe una cantidad muy limitada de trabajos sobre zero-shot Learning en el dominio del procesamiento del lenguaje natural o NLP. En el año 2016 que fue publicado este es el primer trabajo que reporta una solución de zero-shot Learning para la categorización de texto

II-C1. Metodo: En este trabajo se propusieron muchos enfoques de zero-shot Learning en el dominio de la visión computacional por Sandouk Chen (2016), Socher et al. (2013). Sin embargo, existe una cantidad muy limitada de trabajos sobre zero-shot Learning en el dominio del procesamiento del lenguaje natural o NLP. En el año 2016 que fue publicado este es el primer trabajo que reporta una solución de zero-shot Learning para la categorización de texto. La arquitectura que se preenta es una red neuronal de una sola capa en la concatenación de 1. la inserción media de la oración y 2. la inserción de la etiqueta. Está inspirado en arquitecturas superficiales que obtienen buenos puntajes en tareas de clasificación de texto como Joulin et al. (2016). La segunda arquitectura, en lugar de tomar una media de incrustaciones antes de pasarla a la capa de clasificación, intenta modelar la secuencia utilizando un LSTM Hochreiter Urgen Schmidhuber (1997). Nuestra tercera arquitectura LSTM puede considerarse similar a la arquitectura utilizada por Wang et al. (2016) para el análisis de sentimientos basado en aspectos. En lugar del "Término de aspecto", pasamos la inserción de la etiqueta para que se considere relacionada

II-C2. Conclusion: En este trabajo,se preesentaron técnicas y modelos que se pueden utilizar para la clasificación de Zero-Shot learning en textos. Se prueba que los modelos pueden ser mejores que las precisiones de clasificación aleatoria en conjuntos de datos sin ver ni un ejemplo.Se puede decir que esta técnica aprende el concepto de relación entre una oración y una palabra que pueden extenderse más allá de los conjuntos de datos. A traves de esto los niveles de precisión dejan mucho margen para trabajos futuros.

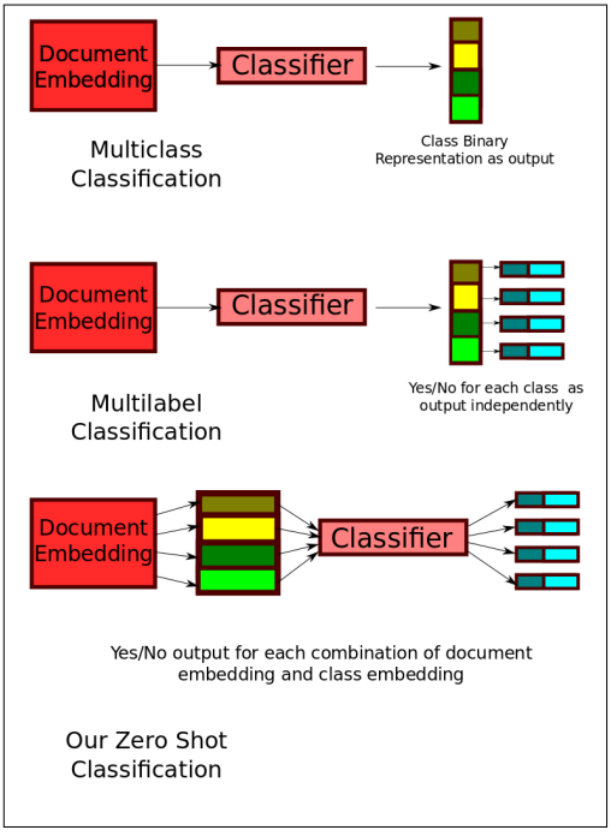

Figura 2. Arquitectura de clasificación multiclase, multietiqueda y clasificación zero-shot learning propuesto

III. Metodología

En esta sección se describen las herramientas y la métodología que se usarán en el presente trabajo

III-A. Metodología de trabajo

Para implementar este trabajo utilizaremos el metodo basado en embedding, para ellos listaremos previamente los 2 tipos de datos(datos de entrenamiento y datos de Zero shot Learning(ZSL)).

No utilizaremos imagenes de ZSL en el entrenamiento ya que no forma parte del modelo, sin embargo si necesitamos representar sus clases como datos.

1. Realizaremos un embedding de imagenes para los datos de entrenamiento por medio de una red convolucional, en este caso una red entrenada.
2. Realizaremos un embedding de clases(de los datos de entrenamiento y de ZSL), para ello usaremos word2vec con el objetivo de que la red aprenda a relacionar una entrada con un vector del espacio word2vec.
3. Realizaremos el embedding de la imagen a clasificar por ZSL, para comparar el vector de caracteristicas de la imagen con todos los vectores de clase que tenemos tanto de entrenamiento como de ZSL por medio de una busqueda vecinos mas cercanos.

III-B. Métricas

En el presente trabajo se evaluará la precision promedio por clase-top-1, metrica para evaluar el rendimiento de Zero shot learning Es decir encontramos la precisión del reconocimiento para cada clase por separado y luego la promediamos entre todas las clases.

Para un conjunto de clases Y con N clases, la precisión promedio por clase-top-1 viene dada por:

$$a_y = \frac{1}{N}\sum_{c=1}^{N}\frac{numerodepredicciones correctas}{numerodemuestras en C}$$

III-C. Pytorch-Transformers

PyTorch-Transformers [**?**] es una biblioteca de modelos pre-entrenados de última generación para el procesamiento del lenguaje natural (NLP), que ahora se llama Transformers y es desarrollado por HuggingFace.

Esta biblioteca contiene implementaciones de PyTorch, pesos de modelos previamente entrenados, scripts de uso y utilidades de conversión para los siguientes modelos: BERT, GPT, GPT-2 (de OpenAI), Transformer-XL, XLNet, XLM.

III-D. SimpleTransformers

SimpleTransformers [**?**] es una libreía construida en base a la biblioteca Pytorch-Trasformers [**?**], que también contiene modelos de arquitectura Transformer pre-entrenados pero cuyo objetivo principal es simplificar la codificación y evaluación de los modelos.

IV. Experimentación y resultados

V. Conclusiones

VI. Trabajos Futuros

Referencias

[1] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In

[2] Sergey Ioffe, Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift (2015). Disponible en https://arxiv.org/abs/1502.03167.

[3] Jimmy Lei Ba, Jamie Ryan Kiros, Geoffrey E.Hinton. Layer Normalization (2016). Disponible en https://arxiv.org/abs/1607.06450.

[4] Kishore Papineni , Salim Roukos , Todd Ward , Wei-jing Zhu (2002). BLEU: a Method for Automatic Evaluation of Machine Translation, pp. 311-318. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL).

[5] An embarrassingly simple approach to zero-shot learning ,Bernardino Romera-Paredes,Philip H. S. Torr Disponible en https://proceedings.mlr.press/v37/romera-paredes15.pdf

[6] PP-ShiTu: A Practical Lightweight Image Recognition System. Disponible en https://arxiv.org/pdf/2111.00775v1.pdf.

[7] Zero-Shot Learning - The Good, the Bad and the Ugly ,Yongqin Xian, Bernt Schiele, Zeynep Akata Disponible en https://arxiv.org/pdf/1712.05972.pdf.