

Zero shot Learning: Clasificación de imágenes

William Guadalupe
FC-UNI
Lima, Perú
wguadalupeq@uni.pe

Victor Chavez
FC-UNI
Lima, Perú
vchavezb@uni.pe

Stefano Olivieri
FC-UNI
Lima, Perú
solivierir@uni.pe

Resumen—

I. INTRODUCCIÓN

I-A. Problemática

Durante las últimas décadas, las máquinas se han vuelto mucho más inteligentes, pero sin un conjunto de datos de entrenamiento debidamente etiquetado de las clases vistas, no pueden distinguir entre dos objetos similares. Sin embargo, los humanos son capaces de identificar aproximadamente 30.000 categorías de objetos básicos. En el aprendizaje automático, esto se considera el problema de Zero-shot Learning (ZSL).

I-B. ¿Por que Zero-Shot Learning?

El desarrollo de modelos de aprendizaje automático pueden realizar funciones predictivas en datos que antes no se habían visto, esto se ha convertido en un área de investigación importante conocida como Zero-Shot Learning. Las personas tenemos la tendencia a ser buenos para reconocer cosas a nuestro alrededor que no hemos vistos antes y Zero-Shot Learning nos ofrece una posible manera para imitar esta capacidad humana.

I-C. Objetivos

I-C1. Objetivos Generales:

-
-

I-C2. Objetivos Específicos:

-
-

I-D. Herramientas y Metodos

I-D1. Herramientas: 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.

I-E. Hipótesis

I-F. Justificación

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.

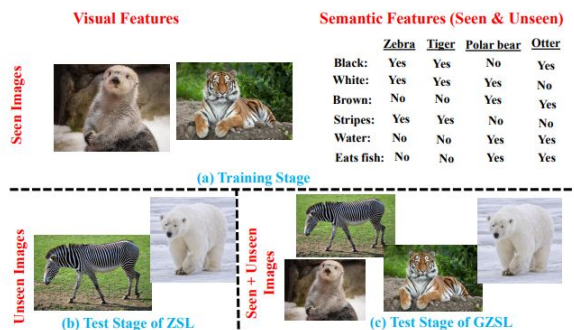


Figura 1. Zero-Shot

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 Uergen 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 presentaron 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 través de esto los niveles de precisión dejan mucho margen para trabajos futuros.

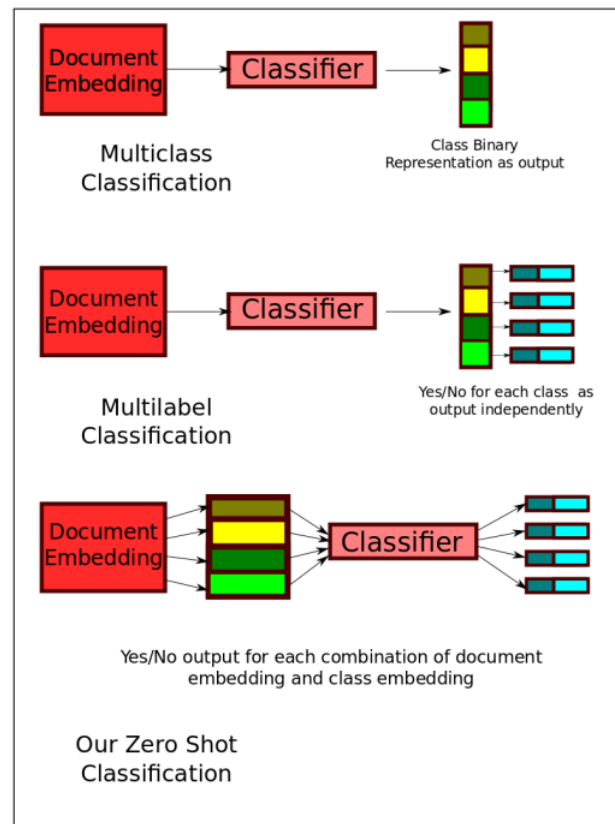


Figura 2. Arquitectura de clasificación multiclase, multietiquetada y clasificación zero-shot learning propuesto

Figura 4. Clases vs Atributos

ResNet-18 es una red neuronal convolucional que tiene 18 capas de profundidad. Puede cargar una versión previamente entrenada de la red entrenada en más de un millón de imágenes desde la base de datos ImageNet. La red previamente entrenada puede clasificar imágenes en 1000 categorías de objetos, como teclado, mouse, bolígrafo y muchos animales.

VI. CONCLUSIÓN

-
-

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] A Review of Generalized Zero-Shot Learning Methods. Disponible en <https://arxiv.org/pdf/2011.08641.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>
- [8] Simple Trnasformers <https://simpletransformers.ai/>
- [9] Pytorch Transformershttps://pytorch.org/hub/huggingface_pytorch-transformers/