Boceto TFG

Hay que reorganizar las cosas, es un boceto

1. INTRODUCCION Y OBJETIVOS

La inteligencia artificial es una herramienta en auge en estos momentos en los que nos encontramos, la mayoría de la población conoce o ha oído hablar de este término tan nombrado que asombra a muchos e inquieta a otros tantos. Pero realmente ¿Qué es la inteligencia artificial?

La inteligencia artificial o IA se considera un campo de la informática especializado o enfocado en crear sistemas que sean capaces de realizar tareas que, normalmente, requieren inteligencia humana como puede ser la percepción, el aprendizaje, el razonamiento u otras actividades. En resumen, buscar que una “máquina” realice tareas o imite la forma de pensar o actuar de un ser humano.

La primera vez que se comenzó a hablar sobre IA fue en 1943 cuando los científicos Warren McCullough y Walter Pitts publicaron el artículo «A Logical Calculus of Ideas Immanent in Nervous Activity» en el que presentan el primer modelo matemñatico para la creación de una red neuronal aunque no fue hasta 1950 cuando dos alumnos de Harvard crearan el primer ordenador de red neuronal y se creara el Test Turing que a dia de hoy se utiliza para valorar las IA.

//

Durante estos años buscaron formalizar matemáticamente el comportamiento de una neurona y de estudiar su forma de computar y procesar la información. Comenzaron con la manipulación de símbolos mediante la cual, la computadora leia uno a uno los símbolos, los consultaba en una tabla y sabia hacia que casilla debía de moverse si izquierda o derecha, en la actualidad este modelo de computación construye la base de la mayoría de lenguajes de programación, toman variables de entrada que se manipulan algebraicamente y devuelven un resultado.

Posteriormente, se propuso un teorema en el que, en lugar de queel autómata manipulara símbolos se crea una red neuronal artificial en la que se iban excitando unas a otras por medio de conexiones sinápticas (nuestros neurotransmisores)

//

La inteligenica artificial comienza alrededor de 1950 como una técnica en la que la computadora puede imitar e comportamiento humano, pasa en los 80 a denominarse machine learning, consiste en técnicas que proporcionan a los ordenadores la abilidad de aprender sin haber sido programados explícitamente para ello. Llegamos al ultimo tramo de la inteligencia artificial que surge en 2010, el Deep learning en el que un subconjunto de aprendizaje automático que hace factible el cálculo de redes neuronales multicapa

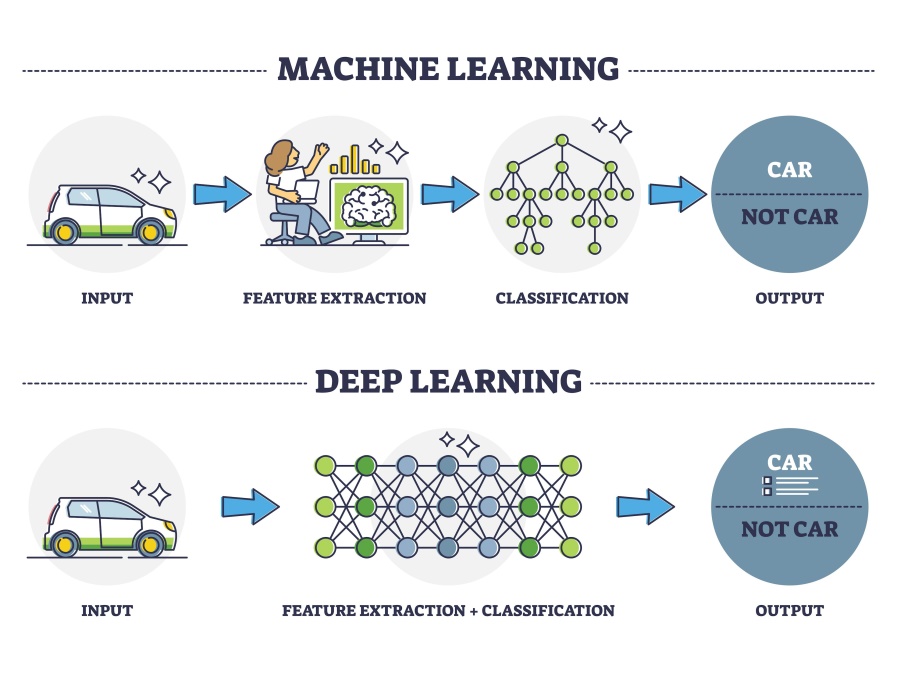
La inteligencia artificial se basa en estudios académicos, logros e investigaciones de diferentes ramas como lo son la probabilidad, la estadística, la ciencia de computadores, las bases de datos, la ciencia cognitiva, neurociencia o patrones de reconocimiento

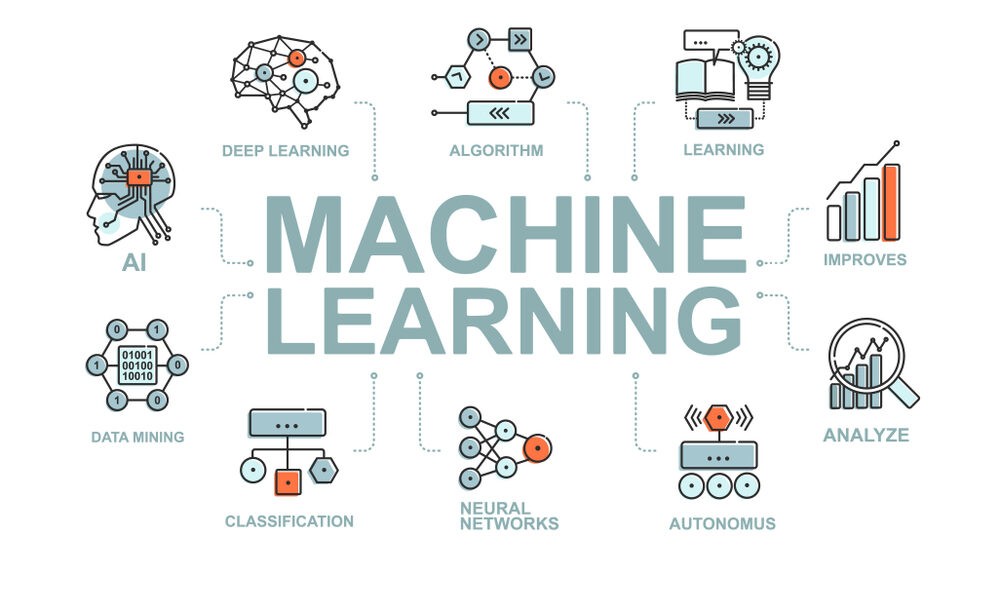
MACHINE LEARNING

En el machine learning lo que buscamos es preparar un dataset para entrenar y extraer las características más significativas de cada elemento del entrenamiento y gracias a estas características luego podemos clasificar nuevos objetos que se asemejen a los ya conocidos

La IA sufrio un paron durante todos estos añis porque no había maquinas suficientespara los procesos de calculo que se necesitaban hacer pero gracias a la incorporación de las tarjetas graficas o GPU

Tarjetas graficas,





Diferencias entre Machine Learning y Deep Learning

Definición y Alcance:

Machine Learning (ML): Es un subcampo de la inteligencia artificial enfocado en desarrollar algoritmos que permiten a las máquinas aprender a partir de datos y hacer predicciones o decisiones sin ser programadas explícitamente. Incluye métodos como la regresión, clasificación, clustering, árboles de decisión, máquinas de soporte vectorial (SVM), y k-vecinos más cercanos.

Deep Learning (DL): Es una subárea de ML que utiliza redes neuronales artificiales con muchas capas para modelar y aprender representaciones de datos con altos niveles de abstracción. Se centra en redes neuronales profundas y sus variantes como las redes neuronales convolucionales (CNN), redes neuronales recurrentes (RNN), redes generativas adversarias (GAN), y modelos de aprendizaje profundo de refuerzo.

Estructura y Algoritmos:

Machine Learning: Utiliza algoritmos más sencillos y menos complejos en términos de estructura, como la regresión lineal y los árboles de decisión. Requiere que los ingenieros de datos seleccionen y diseñen manualmente las características importantes del modelo. Los modelos suelen ser menos complejos y no requieren grandes cantidades de datos.

Deep Learning: Emplea redes neuronales profundas con muchas capas que pueden aprender automáticamente representaciones jerárquicas de los datos. No necesita intervención manual significativa para el diseño de características. Requiere modelos muy grandes y grandes cantidades de datos etiquetados para entrenarse eficazmente.

Requisitos de Datos y Computación:

Machine Learning: Puede funcionar con conjuntos de datos más pequeños y estructuras menos complejas. Es menos intensivo computacionalmente y puede ejecutarse en máquinas normales sin necesidad de hardware especializado.

Deep Learning: Necesita grandes cantidades de datos para ser efectivo y es altamente intensivo en computación, requiriendo hardware especializado como GPUs para entrenar los modelos eficientemente.

Aplicaciones Típicas:

Machine Learning: Se aplica en detección de fraudes, análisis de sentimientos, recomendaciones de productos y modelos predictivos de negocios. Es útil para tareas donde los patrones en los datos no son extremadamente complejos.

Deep Learning: Se utiliza en reconocimiento de voz, visión por computadora, traducción automática, conducción autónoma y diagnóstico médico. Es ideal para tareas donde los datos tienen estructuras complejas y se necesita una alta precisión.

Resumen de las Diferencias

| Característica | Machine Learning | Deep Learning |
| --- | --- | --- |
| Definición | Subcampo de IA para algoritmos de aprendizaje | Subcampo de ML utilizando redes neuronales profundas |
| Algoritmos | Más simples (regresión, SVM, árboles de decisión) | Complejos (CNN, RNN, GAN) |
| Diseño de Características | Manual | Automático |
| Tamaño del Modelo | Menos complejo | Muy complejo |
| Requisitos de Datos | Menos datos | Gran cantidad de datos |
| Computación | Menos intensivo | Altamente intensivo |
| Aplicaciones | Tareas tradicionales (detección de fraudes, recomendaciones) | Tareas complejas (reconocimiento de voz, visión por computadora) |

Esta tabla y descripción muestran las diferencias clave entre Machine Learning y Deep Learning en términos de definición, estructura, requisitos y aplicaciones.

lo que mas usamos librerías de numpy para cálculos matemáticos y librerías de pandas para la manipulación y análisis de datos flexibles y de alto rendimiento

La IA se basa en la rama de las matemáticas relacionadas con la probabilidad y la estadística, es principalmente eso. Concretamente vamos a referirnos al campo de la estadística inferencial ya que Utiliza métodos que permiten generalizar los resultados obtenidos en una muestra al conjunto de la población, proporcionando estimaciones y pruebas de hipótesis

Usos de la probabilidad y la estadística relacionado con la IA

Procesamiento de Lenguaje Natural (NLP): Usa probabilidades para predecir la probabilidad de secuencias de palabras.

Aprendizaje Automático (Machine Learning): Estadística inferencial para evaluación y validación de modelos

Para toda la parte de Deep learning y concretamente en este estudio vamos a utilizar TensorFlow, un framework para programar aplicaciones de IA con Python compatible tanto para CPU como para GPU

En cuanto a la arquitectura para las redes neuronales del Deep learning vamos a utilizar capas convolucionales, densas y de dropout. La combinación de capas no permiteir haciendo el análisis de las imágenes y evitar el sobreajuste entre ellas. Ademas todo esto se ve afectado por las redes recurrentes que mantienen la unión entre las neuronas asi como la entrada y salida de la información cuando tratamos datos secuenciales

REDES CONVOLUCIONALES: hablando sobre Deep learning

Encuanto a las redes neuronales convolucionales (CNN) es una red neuronal profundaque se especializa en clasificar imágenes

El objetivo final del trabajo es que el sistema sea capaz de ofrecer una descripción de una imagen. Para ello voy a utilizar un sistema de Deep learning, que consiste en el entrenamiento de un algoritmo que es capaz de reconocer las características de las 8000 imágenes que contiene el dataset y aprender su contenido para posteriormente ser capaz de describir una imagen nueva que se le proporcione al algoritmo y que no pertenezcan al conjunto de datos del entrenamiento.

1. ESPECIFICACIÓN Y REQUISITOS

En cuanto a la especificación del trabajo, es necesario una gran cantidad de imágenes provenientes de un dataset en la red o de una base de datos. Todas esas imágenes deben tener descripción e imagen, en este caso cada imagen contiene 5 descripciones válidas

En cuanto a requisitos funcionales se necesita procesar las imágenes detectar y clasificar los objetos que contienen para asi crear los nuevos textos descriptivos y el audio

Para ello es necesario tener mucho tiempo de funcionamiento de la máquina para los entrenamientos y el procesamiento, es necesaria una GPU y suficiente memoria ram o en su defecto contar con un servicio online que te lo pueda proporcionar

Si queremos guardar toda esta información en una bbdd lo que haremos será contratar servicios de Azure, Google Cloud Platform o un servicio en el que podamos almacenar imagen y texto sin necesidad de crear relaciones entre ellos como podría ser en MongoDB.

1. Planificacion temporal y evaluación de costes

|  |  |
| --- | --- |
| Establecer las partes del proceso que se tienen que realizar | 1 semana |
| Comenzar a programar cada una de las partes necesarias para la creación del código que será capaz de generar las descripciones y analizar las imagenes | 5 semanas |
| Mejorar la calidad de esas descripciones y añadir fotos externas a las utilizadas para el entrenamiento | 2 semanas |
| Implementacion y análisis de las posibles mejoras a largo plazo | 2 semanas |

Costes??

1. Tecnologías utilizadas

Desarrollar cada uno con la información de la introducción

Modelos que vamos a utilizar para nuestra IA

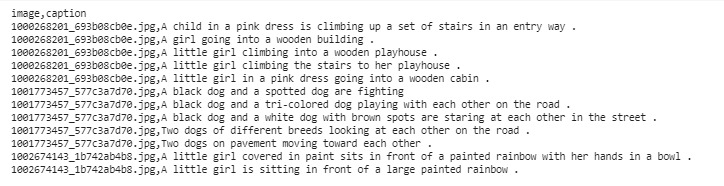
Redes neuronales convolucionales CNN para el reconocimiento y clasificación de objetos en imágenes, modelos de lenguaje natural NLP para la generación de descripciones

En cuanto a los entornosde desarrollo por ahora se van a utilizar Jupyter para manejar toda la parte de back con Python y, en la implementación de una parte front utilizaremos html con css que crearemos en jupyther o vscode

Para el control del proceso y versiones lo realizaremos mediante git

5. DESARROLLO E IMPLEMENTACIÓN

Parto del conjunto de datos que nos proporciona Kaggle. Kaggle es una comunidad de inteligencia artificial y machine learning que proporciona recursos, modelos entrenados o conjuntos de datos entre otras cosas, a estudiantes, desarrolladores e investigadores que quieren trabajar en esta parte de la inteligenica artificial.

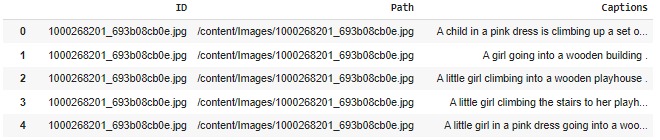
El conjunto de datos que he elegido consta de 8000 fotos, el nombre es flickr8k (<https://www.kaggle.com/datasets/adityajn105/flickr8k/data>), además, viene acompañado de un fichero que contiene las descripciones de esas fotos. El fichero se llama caption.txt, en el encontramos el nombre del archivo jpg que contiene la imagen, una descripción de la imagen y la ruta física donde se aloja el archivo jpg. Por cada imagen hay 5 descripciones en el fichero 

Para trabajar con este dataset montare una unidad en drive para descargar toda la información que vamos a usar.

(Explicar problemas de almacenamiento en bbdd)

Esta primera versión se corresponde con el tratamiento que hacemos de las 8000 fotos y de sus 40000 descripciones . Para ello utilizamos dos algoritmos de entrenamiento, por un lado tratamiento de imágenes y por otro el texto que describe las imágenes.

En primer lugar hay que tratar el conjunto de datos y construir un dataset que tenga la información que nos interesa, en este caso, construimos un dataset firmado por un id, una ruta y la descripción de la foto.



Primero tratamos el texto, hay que hacer un trabajo de limpieza de las descricpiones, comprobar si hay minúsculas, mayúsculas, dividir todo en palabras… y creamos una lista que contenga todas las palabras que hemos obtenido para luego poder usarlas en el proceso de crear la nueva descripción de la foto, además de usarlo en el proceso de aprendizaje

A partir de ahí, una vez tratado el texto, comenzamos con el tratamiento de las imágenes. Para el tratamiento de las imágenes, utilizamos un modelo pre-entrenado de imágenes porque los medios físicos que tengo para trabajr no son suficientes para desarrollar un model entrenado por mi porque es necesario disponer de una maquina muy potente, con GPU potente. Utilizamos el modelo pre-entrenado VGG16 que sirve para procesar imágenes. Para usar este modelo es necesario que todas las imágenes tengan unas determinadas características, para ello voy a crear unas funciones que hacen que todas las fotos del modelo cumplan estas características. Todas als imágenes tienen que estar 224x224 pixeles y cada pixel tiene que tener un valor entre 0 y 1 (normalización : crear un tensor únicamente de 0 y 1 para trabajar con ello en el aprendizaje)

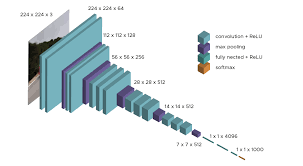
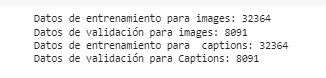


Grafico de la arquitectura del modelo VGG16 en deep learning

Una vez hecho el tratamiento de las imágenes se divide el conjunto de datos en dos partes, una parte para entrenar y otra para probar. En este caso he elegido un conjunto de entrenamiento del 80% de las fotos y para las pruebas el 20% restante 

El trabajo del entrenamiento de las imágenes lo realiza de forma automática este modelo (reconocer lo que contiene cada foto y su descripción asociada)

Ahora pasamos a la parte del texto, una vez entrenadas las imágenes pasamos al entrenamiento de las descripciones de cada una de esas imágenes. Lo que intento hacer en esta parte es que el sistema sea capaz de predecir una descripción de cualquier imagen que se le indique y que pertenezca al conjunto de entrenamiento. Para ello construyo un modelo de procesamiento del lenguaje natural (NLP) que consiste en un codificador y decodificador de palabras.

Este algoritmo aprenderá de todas las palabras que componen las descripciones de las fotos en un proceso de entrenamiento, realiza la codificación de las imágenes y la decodificación de las mismas para generar descripciones textuales.

El proceso de decodificación, o de generación de descripciones pertenece a lo que se denomina IA generativa. La IA generativa es un tipo de inteligenica artificial que puede crear ideas y contenidos nuevos, en este caso generara nuevas descripciones para una foto que pertenezca al conjunto de datos que hemos usado para el entrenamiento y debe ser funcional para fotos externas.

El proceso de entrenamiento del modelo de texto se genera con varias épocas, una época es el tiempo que el algoritmo tarda en procesar todas las fotos y cada una de sus descripciones. Por cada época añadida el proceso va mejorando el aprendizaje para afinar la descripción que hace de cada una de las fotos. He realizado dos entrenamientos, uno con 5 epocas, otro con 10 y el objetivo final serian 20 epocas que es el tiempo máximo de uso que me permite google colab. En cada entrenamiento veremos como mejora la descripción generada de fotos que pertenecen al conjunto de entrenamiento y de una foto que no pertenece al conjunto.

1. Líneas futuras

DESARROLLAR y las conclusiones

Este código es el backend de la aplicación, donde se realiza el trabajo. Mi idea sería construir un fornt que permitiera subir una imagen y que ofrezca la descripción de la misma.

Para ello tengo que hacer una aplicación o web en kotlin o html, que permita al usuario subir una foto, que esa foto se guarde en el drive y en el back, tratarla, normalizarla, generar la descripción y el audio y devolverlo al front para el usuario.

Bibliografia

Apuntes curso IA Samsung innovation Campus

<https://es.mathworks.com/discovery/artificial-intelligence.html>