Sistemas de aprendizaje automático

10/02/2023

EL profesor se llama cesar, no exige presencialidad

Hay otro profesor que se llama Andreas, especialista en IA

Introducción

<https://www.cesguiro.es/doku.php>

Katas para IA: kaggle

/\* Tarde \*/

Andreas Lloyd

16/10/2023

La base de la IA es todo matemáticas y estadística, sobre todo el machine learning.

Definir IA es complicado, está presente en todos lados y normalmente la sociedad cuando piensa en IA lo hace en Inteligencia Artificial Fuerte.

Por ejemplo los sistemas de recomendaciones son IA, utilizando sistemas de aprendizaje no supervisado (como clusterización). Sin embargo, esto al final es un sesgo, porque siempre te recomienda dentro de la zona de confort y es más difícil abrir a nuevos estímulos.

Los mayores problemas de la IA son los datos y los sesgos, los cuales metemos sin querer en muchas ocasiones.

El procesamiento de imágenes y audios si que es una herramienta muy practica.

¿Qué es IA?

Es complicado definir el concepto de IA. 2 características

* Autonomía:
* Adaptabilidad: ir aprendiendo de los datos

IA débil vs IA fuerte

La gente que no sabe mucho entiende IA como hacer un robot “Terminator”. Esto no es lo que realmente se está trabajando

Débil vs Fuerte

* La IA débil existe, la fuerte no (solo en películas)
* Orientado a problemas concretos/Resuelve problemas abiertos (
* Reactivo/Proactivo
* Sin flexibilidad/ Con flexibilidad
* Programada por un humano/Se autoprograma
* Pocas redes neunorales/ Muchas redes neuronales a veces en conflicto
* No razona, solo computa/Imita el comportamiento humano
* Aprende de ejmplos similares/ Aprenden como humanos
* Tareas repetitivas/Aprenden nuevas tareas
* No se puede salir de su marco de trabjo/Adaptabilidad a nuevos escenarios

Aprendizaje automático

Machine learning modelos que aprendizajen por si mismos, dentro del machine learning está las redes neuronales y dentro de esta el Deep learning

Tipos de machine learning

* Aprendizaje supervisado: en la fase de entrenamiento los datos están etiquetados. Ejemplo: le doy fotos y le digo que es (Gato o perro). Una vez en producción puede predecir nuevos resultados en base a datos que nunca ha recibido.
  + Regresión lineal: existe un numero infinito de posibilidades. Sacar una función que en función de unas variables te predice un valor. Ejemplo: altura de los padres y altura de los hijos.
  + Clasificación: diferencia entre una serie de posibilidades concretas (perro o gato). Por ejemplo: detectar si un correo es spam
* Aprendizaje no supervisado: datos no catalogados. El programa debe agrupar (clustering) o reducción de dimensionalidad (eliminar características no útiles)
* Aprendizaje semi-supervisado: solo tiene unos pocos datos etiquetados, utiliza ambas técnicas para etiquetar y agrupar los daots
* Aprendizaje por refuerzo: a través de un sistema de castigo y recompensa es capaz de aprender por si mismo. Por ejemplo: un laberinto

Aprendizaje por lotes/online

* Aprendizaje offline: tienes el modelo, le pasas todos los datos y lo pones en producción, si lo quieres re-educar tienes que volver a hacer todo el proceso
* Aprendizaje por lotes:
* Aprendizaje online: tiene un parámetro denominado tasa de aprendizaje que indica la importancia que tiene los nuevos datos en función a los antiguos

Aprendizaje basado en instancias/basado en modelos

* Aprendizaje basado en instancias: predice las nuevas instancias en función a la similitud con los datos más cercanos
* Aprendizaje en modelo: genera zonas diferentes e incluye las nuevas instancias en una de estas zonas

Problemas del machine learning

* Tener pocos datos: sobreajustando los datos, en la fase de entramiento acierta todos los datos, pero en producción es incapaz de predecir. No encuentra patrones, aprende de memoria los datos
* Datos de entrenamientos no representativos: datos sesgados de forma intencional o no. Ejemplo de reconocimientos facial con hombre de piel oscura.

Ejemplo de problemas

* Sobreajustar los datos: no generaliza bien
* Subajustar los datos: modelo demasiado simple

**Los datos tienen que ser representativos**

**Es necesario tener una cantidad suficiente de datos y de características para hacer un modelo bueno**

Utilizar los 80% de los datos para entrenar y el otro 20% de datos para validar, estos nunca se deben mirar antes de utilizarlos.

También se pueden separar los datos de entrenamiento para ir ajustando el modelo en función del aprendizaje.

Una vez está el modelo, se vuelve a entrenar con TODOS los datos para entrenarlo con la totalidad de datos, sin dejar ninguno

Redes neuronales

**Perceptrón**

**Tiene una serie de entradas, a cada entrada se le asigna un peso y produce una función lineal. Para romper la linealindad se utilizan funciones de activación.**

Deep learning

El Deep leaning es un caso particular de las redes neuronales, consiste en conectar varias capas entre sí. Ejemplo: si tienes una foto, sacas cada uno sus pixeles y eso son las entradas que tiene el modelo. Las salidas dependen del resultado que quieras obtener.

Backpropagation

Entrenar hacia atrás con datos etiquetados, en función del error ajusta las neuronas

Tipos de redes neuronales

Convulocionales: Para imágenes, cada una de las capas detectan diversos aspectos de la imagen (formas, etc…)

Recurrentes: procesamiento del lenguaje natural, procesamiento del texto. Necesitan recordar información anterior. Por ejemplo: para poner en contexto frases o textos

Generativa adversarias: hay 2 redes diferentes, un generador y un discriminador. Ejemplo: un falsificador de billetes y un policía. El falsificador sabe lo que piensa el policía. A medida que el falsificador sabe por que se rechazan los billetes los va mejorando hasta que hace billetes que son “reales”

/\* Tarde con Andreas Lloyd \*/

Importante saber inglés.

Primer ejercicio

Contexto:

Objetivo: comprar un piso en Iowa a muy buen precio. Ojo: el objetivo no es entrenar el modelo. Ganar confianza sobre los precios para poder juzgar si estamos comprando a algo a buen precio

Metodología: crear un modelo de regresión para poder el precio de regresión.

Datos: hay un histórico de ventas, con una descripción en .txt

**¿Qué se espera de nosotros?**

Abrir el notebook, seguir las instrucciones y responder a las preguntas. En notebook te guía por la metodología típica de un modelo de IA. No siempre se aplican todos los pasos de la misma forma, pero es una estructura básica.

A las preguntas no solo se responde con texto, la metodología y la justificación de la respuesta también forman parte de la respuesta.

Los notebooks mezclan código con markdown

Se utilizan principalmente las librerías de Pandas (manejar datos) y scikit-learn (machine learning, es muy flexible)

Hay 7 pasos que generalmente se siguen a la hora de crear un modelo

El análisis puede llevar años

Es fundamental que los datos estén bien

Las fechas y horas usualmente no son buenas variables porque los modelos pueden utilizarlas para afinar. Sin embargo, fechas más genéricas como el mes si que puede ser de utilidad, o ciertos rangos de horas también pueden ser significativos (los usuarios registrados a las 3am son menos serios)