

ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS

MACHINE LEARNING ALGORITHMS FOR DATA ANALYSIS AND PREDICTION

Lilian Judith Sandoval.

Licenciada en Administración de Empresas, con diplomado en Tecnologías de Software.

Docente de la Escuela de Ingeniería en Computación

ITCA-FEPADE Sede Central, Santa Tecla.

jose.peraza@itca.edu.sv

Recibido: 16/04/2018 - Aceptado: 19/07/2018

Resumen

Actualmente podemos ver que hemos entrado en una nueva era de información en la que las compañías conocen de antemano nuestras preferencias y, de acuerdo con nuestro comportamiento en la red, pueden predecir qué productos preferiremos antes de su lanzamiento. Usan nuestra información para generar nuevas campañas de marketing con mayor seguridad de que los productos tendrán una mayor aceptación y al mismo tiempo reducirán los riesgos. Todo este conocimiento es proporcionado gracias a la Ciencia de Datos mediante la técnica del Machine Learning. Trabajan con cantidades gigantescas de información, las que, con seguridad, establecerán un comportamiento en los datos con el tiempo proporcionando patrones que con mucha probabilidad se seguirán dando en el futuro. Es así como se genera la predicción de los datos.

Abstract

At present we can see that we have entered into a new age of information in which companies know in advance our preferences and, according to our behavior in the network, they can predict what products we will like before them to released those products. They use our information to generate new marketing campaigns with improve certainty that the products will have greater acceptance and at the same time reduce any possible risk. All this knowledge is provided by Data Science, through the Machine Learning approach. They work with huge amounts of information, with enough time, all this data will certainly establish a behavior, providing patterns with great probability to keep occurring in the future. That's how data prediction is generated.

Palabras clave

Inteligencia artificial, inteligencia de negocios, Big Data, bases de datos.

Keyword

Artificial intelligence, business intelligence, Big Data, databases.

Introducción

Con la cantidad de información que es generada día con día en Internet, ya sea por redes sociales, transacciones comerciales, datos emitidos por distintos dispositivos, etc. existen procesos que aprovechan toda esa información y en lugar de conservarla como data almacenada que solo está ocupando mucho espacio en los servidores, se sigue utilizando para hacer análisis de comportamientos y de algún modo identificar tendencias futuras. Muchas veces se reúne tanta información, que es posible conocer con anticipación

y de forma segura cuál será el comportamiento de un grupo de personas o equipos electrónicos en un futuro.

Conoceremos sobre la técnica del Machine Learning, elemento fundamental de la Ciencia de Datos, los métodos que utiliza para realizar las predicciones de datos y su presentación.

Inteligencia Artificial

Los dispositivos que cuentan con inteligencia artificial pueden ejecutar distintos procesos análogos al comportamiento humano, como la devolución de una respuesta por cada entrada (similar a los reflejos de los seres vivos), la búsqueda de un estado entre todos los posibles según una acción o la resolución de problema mediante una lógica formal.

Cuando se otorga a estos dispositivos la habilidad de aprender y de discernir, se les convierte en entidades que rozan las capacidades de un superhombre, dado que alcanzan velocidades de procesamiento imposibles para los humanos y no necesitan descansar para funcionar, entre otras ventajas que los ubican por sobre los seres vivos en este contexto [1].

Machine Learning

Es una rama de la Inteligencia Artificial que se encarga de generar algoritmos que tienen la capacidad de aprender y no tener que programarlos de manera explícita. El desarrollador no tendrá que sentarse a programar por horas tomando en cuenta todos los escenarios posibles ni todas las excepciones posibles. Lo único que hay que hacer es alimentar el algoritmo con un volumen gigantesco de datos para que el algoritmo aprenda y sepa qué hacer en cada uno de estos casos.

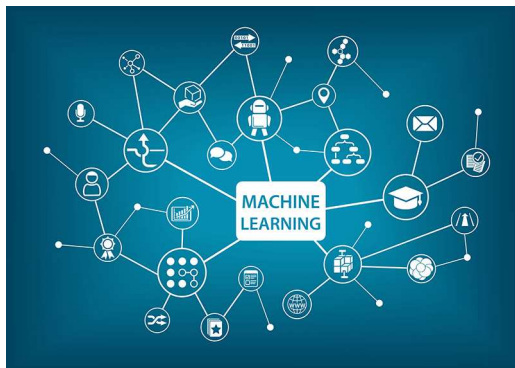


Fig. 1. Fuentes de datos del Machine Learning

Hay dos tipos de aprendizajes: el supervisado y el no supervisado.

A) Aprendizaje supervisado

Es cuando entrenamos un algoritmo de Machine Learning dándole las preguntas (características) y las respuestas (etiquetas). Así en un futuro el algoritmo pueda hacer una predicción conociendo las características.

En este tipo de aprendizaje hay dos algoritmos (entrenamientos): **el de clasificación y el de regresión.**

- **Algoritmo de clasificación:** esperamos que el algoritmo nos diga a qué grupo pertenece el elemento en estudio. El algoritmo encuentra patrones en los datos que le damos y los clasifica en grupos. Luego compara los nuevos datos y los ubica en uno de los grupos y es así como puede predecir de qué se trata.

La variable por predecir es un conjunto de estados discretos o categóricos. Pueden ser:

- *Binaria:* {Sí, No}, {Azul, Rojo}, {Fuga, No Fuga}, etc.
- *Múltiple:* Comprará {Producto1, Producto2...}, etc.
- *Ordenada:* Riesgo {Bajo, Medio, Alto}, etc.

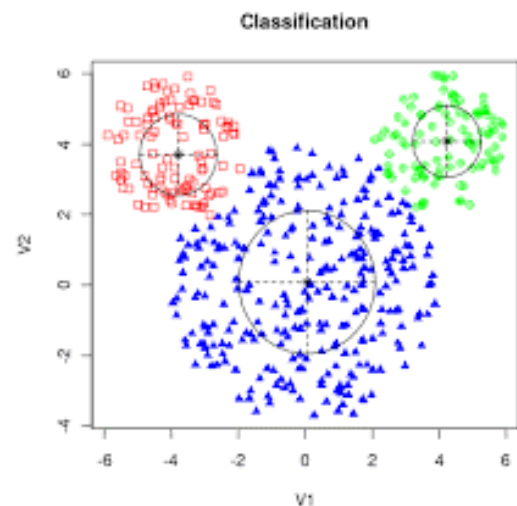


Fig. 2. Gráfico de un algoritmo de clasificación

- **Algoritmo de regresión:** en este método lo que se espera es un número. No lo ubica en un grupo, sino que devuelve un valor específico.

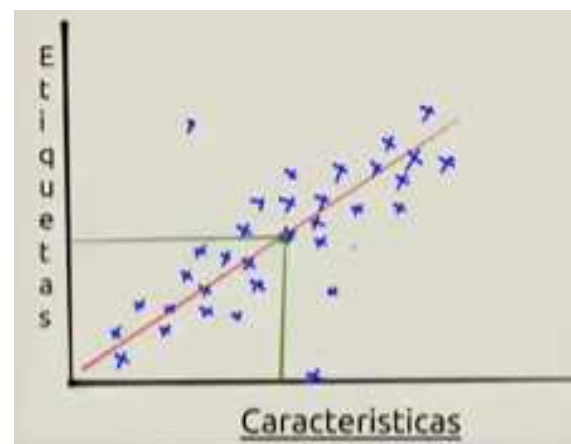


Fig. 3. Gráfico de un algoritmo de regresión

Por ejemplo, el precio de una casa. El algoritmo tiene el precio de diferentes casas, pequeñas, grandes, en el campo, en la ciudad, etc. y por medio de un gráfico de dispersión, puede predecir el precio correcto de una casa en consulta.



Fig. 4. Ejemplo de gráfico de dispersión en un algoritmo de regresión

B) Aprendizaje no supervisado

Aquí solo le damos las características al algoritmo, nunca las etiquetas. Queremos que nos agrupe los datos que le dimos según sus características. El algoritmo solo sabe que como los datos comparten ciertas características, de esa forma asume que pueda que pertenezcan al mismo grupo.

Modelos de Machine Learning

Los algoritmos de Machine Learning, se pueden agrupar en tres modelos:

1) Modelos lineales

Estos tratan de encontrar una línea que se “ajuste” bien a la nube de puntos que se disponen. Aquí destacan desde modelos muy conocidos y usados como la regresión lineal (también conocida como la regresión de mínimos cuadrados), la logística (adaptación de la lineal a problemas de clasificación -cuando son variables discretas o categóricas-). Estos dos modelos tienen el problema del “overfit”, esto significa que se ajustan “demasiado” a los datos disponibles, con el riesgo que esto tiene para nuevos datos que pudieran llegar. Al ser modelos relativamente simples, no ofrecen resultados muy buenos para comportamientos más complicados.

2) Modelos de árbol

Son modelos precisos, estables y más sencillos de interpretar básicamente porque construyen unas reglas de decisión que se pueden representar como un árbol. A diferencia de los modelos lineales, pueden representar relaciones no lineales para resolver problemas. En estos modelos, destacan los árboles de decisión y los random forest (una media de árboles de decisión). Al ser más precisos y elaborados, obviamente ganamos en capacidad predictiva, pero perdemos en rendimiento.

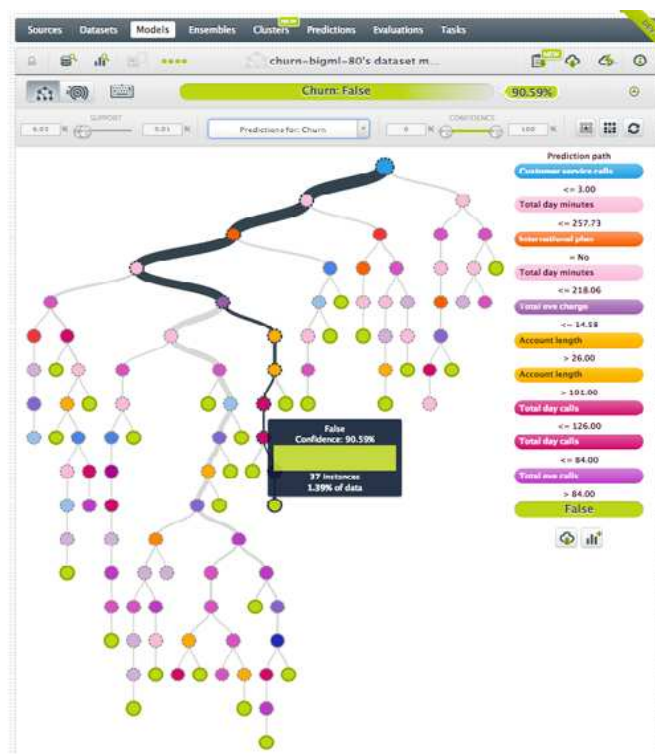


Fig. 5. Gráfico de un Modelo de Árbol

3) Redes neuronales

Las redes artificiales de neuronas tratan, en cierto modo, de replicar el comportamiento del cerebro, donde tenemos millones de neuronas que se interconectan en red para enviarse mensajes unas a otras. Esta réplica del funcionamiento del cerebro humano es uno de los “modelos de moda” por las habilidades cognitivas de razonamiento que adquieren. El reconocimiento de imágenes o vídeos, por ejemplo, es un mecanismo complejo y una red neuronal es lo mejor para realizarlo. El problema, como ocurre con el cerebro humano, es que son lentas de entrenar y necesitan mucha capacidad de cómputo. Quizás sea uno de los modelos que más ha ganado con la “revolución de los datos” [2].

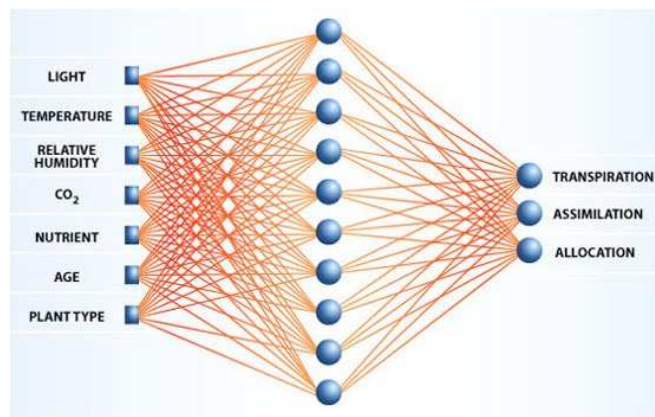


Fig. 6. Gráfico de Redes Neuronales

Fases de desarrollo

A) Fase de entrenamiento

En esta fase se tiene una cantidad enorme de datos, de la cual se separa una parte para entrenar al algoritmo y darle toda esta información para que encuentre los patrones necesarios y después pueda hacer predicciones.

B) Fase de prueba

El resto de los datos que quedan, se van a usar para hacer las pruebas. Así le podemos hacer preguntas al algoritmo y evaluar si las respuestas están bien o mal, y saber si está aprendiendo o no. Si vemos que no coinciden los datos, tendremos que agregar más datos o cambiar el método que estamos utilizando. Pero si se observa que hay entre un 80% a 90% de respuestas correctas, podemos decir que hay un buen grado de aprendizaje y poder utilizar ese algoritmo.

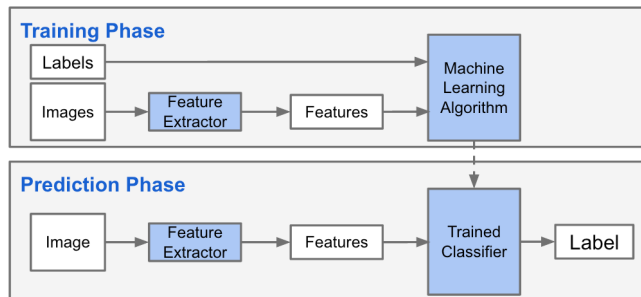


Fig. 7. Fases de Machine Learning.

Ámbitos de la aplicación

Productos que utilizan algoritmos de Machine Learning

- Vehículos no tripulados que se conducen solos.
- Brazo robótico que juega ajedrez.
- Reconocimiento facial de Facebook para identificar contactos.
- Microsoft Cortana, asistente personal inteligente para diferentes dispositivos.
- Motores de búsqueda que ofrecen información de acuerdo a las preferencias de los usuarios.
- Machine Translation usado por el traductor de Google, que reconoce palabras en más de 100 idiomas humanos.
- Google Trends, son las tendencias de búsquedas en Google
- Google N Gram Viewer, indexa libros que tiene Google escaneados y sus términos gramaticales.

- Siri, que convierte conversaciones habladas a texto (STT - Speech To Text) [3].



Fig. 8. Brazo robot que utiliza Machine Learning.

Procesos que hacen uso de Machine Learning.

- Detectar fraudes en transacciones bancarias.
- Detectar intrusiones en una red de comunicaciones de datos.
- Predecir fallos en equipos tecnológicos.
- Prever qué proyectos serán más rentables el próximo año y con un menor riesgo.
- Seleccionar clientes potenciales basándose en comportamientos en las redes sociales, interacciones en la web, etc.
- Predecir el tráfico urbano y dar rutas alternativas.
- Conocer anticipadamente qué partido político ganará las próximas elecciones analizando los comentarios de los usuarios en las redes sociales
- Saber cuál es el mejor momento para publicar tuits, actualizaciones de Facebook o enviar newsletters.
- Prevenir la deserción de clientes en una empresa de telefonía.
- Predecir las ventas de los años siguientes analizando comportamiento actual de los clientes.
- Conocer las preferencias de los clientes a través de sus operaciones en la red.
- Hacer prediagnósticos médicos basados en síntomas del paciente.
- Cambiar el comportamiento de una App móvil para adaptarse a las costumbres y necesidades de cada usuario.

Diferencias entre el Machine Learning y el Business Intelligence

Cualquier tipo de aplicación de Business Intelligence, primeramente, recoge los datos en bruto desde la base de datos transaccional, que es donde se registran todas las operaciones del negocio día con día. Una vez almacenados, los ingenieros de datos utilizan lo que se denomina herramientas de ETL (Extraer, Transformar y Cargar) para manipular, transformar y clasificar los datos en una base de datos estructurada, conocida como Data Warehouse. Luego, los analistas de negocio utilizan técnicas de visualización de datos para explorar los datos almacenados en los Data Warehouse. Con este tipo de herramientas crean paneles visuales (o dashboards) para hacer accesible la información a perfiles de negocio no especialistas en datos. Los paneles ayudan a analizar y entender los resultados en el pasado y sirven para adaptar la estrategia futura que mejore los indicadores clave de negocio.

El Machine Learning, en cambio, es una técnica que permite detectar patrones “a bajo nivel” en miles de datos individuales. El desarrollo de aplicaciones predictivas es una de las potencias destacables de esta técnica, ya que facilita la automatización de procesos, la toma de decisiones y el continuo aprendizaje basado en datos. Además, se trata de sistemas que aprenden automáticamente con el tiempo, se integran en el desarrollo de la compañía y se adaptan a los cambios de entorno cuando se les alimenta de forma constante con nuevos datos.

En un principio podría parecer poca la diferencia, ya que el Machine Learning también usa los datos para trabajar,

utiliza herramientas de ETL para acceder a ellos y su propósito principal es mejorar los objetivos de negocio de las compañías en un futuro.

El mecanismo que lo hace diferente es la detección de patrones entre millones de datos (Big Data). Esta es una diferencia importante respecto a la inteligencia de negocios tradicional, a la que podríamos añadir estos tres aspectos:

- 1) Frente al uso de datos agregados, el Machine Learning utiliza datos individuales con características definitorias de cada una de las instancias. De esta forma se pueden usar miles de variables para detectar los patrones.
- 2) En lugar de basarse en una analítica descriptiva, Machine Learning ofrece una analítica predictiva. Es decir, no solo hace una valoración de lo que ha pasado y extrapola tendencias generales, sino que hace predicciones individualizadas en el que los detalles y matices definen los comportamientos del futuro.
- 3) Los paneles de visualizaciones o dashboards se sustituyen por aplicaciones predictivas. Estamos hablando de uno de los mayores potenciales del Machine Learning: los algoritmos predictivos aprenden automáticamente de los datos y sus modelos se pueden integrar en aplicaciones para dotarlas de capacidades predictivas. Los modelos se reentrenan periódicamente para que aprendan automáticamente de nuevos datos. [4]

Referencias

[1] «Definición de inteligencia artificial - Qué es, Significado y Concepto». [En línea]. Disponible en: <https://definicion.de/inteligencia-artificial/>. [Accedido: 16-mar-2018]

[2] Rayón, «Guía para comenzar con algoritmos de Machine Learning», Deusto Data (blog), 2017. [En línea]. Disponible en: <https://blogs.deusto.es/bigdata/guia-para-comenzar-con-algoritmos-de-machine-learning/>. [Accedido: 16-mar-2018]

[3] A., Conchas, «8 aplicaciones de Machine Learning», 2017. [En línea]. Disponible en: <https://www.inbest.cloud/comunidad/8-aplicaciones-de-machine-learning>. [Accedido: 16-mar-2018]

[4] «Diferencias entre Business Intelligence y Machine Learning», 2017. [En línea]. Disponible en: <https://cleverdata.io/diferencias-bi-machine-learning/>. [Accedido: 16-mar-2018]