

Clarificando conceptos en el ámbito de la ciencia de datos y la inteligencia artificial

La importancia del lenguaje

Toda actividad innovadora tiene la imperiosa necesidad de crear una terminología que permita describir los nuevos conceptos en los que se adentra. En ocasiones los conceptos son totalmente nuevos, en otras son conceptos existentes pero que se trasladan a un nuevo contexto. En ambos casos, se crean palabras que buscan dar un sentido diferencial a esas ideas.

Es normal, además, que existan grupos de trabajo en paralelo desarrollando la actividad, lo que genera a su vez un intercambio adicional que va matizando o reconstruyendo esos conceptos.

En el caso del big data, la inteligencia artificial y la ciencia de datos en general, a pesar de su notable avance, son ciencias relativamente jóvenes y todavía existen fronteras difusas en muchos de sus conceptos.

Por ello, creemos imprescindible comenzar cualquier labor divulgativa especificando el uso que se le dará a cada una de las palabras. No asumimos que el vocabulario que aquí especificamos sea universalmente aceptado, aunque queremos poner nuestro granito de arena en el proceso. Lo que sí buscamos es clarificar el uso que vamos a dar a cada uno de los conceptos, de forma que nuestros interlocutores tengan claro de qué estamos hablando en cada caso. Lo importante no es la palabra en sí misma, lo importante es que cuando digamos una palabra, quién nos escuche o nos lea tenga claro a qué nos referimos -aunque sea tan rara como "exalutenia", vocable que ni siquiera hemos encontrado en Google-.

Una visión inicial de alto nivel

A la hora de abordar este ámbito, creemos que es más eficiente comenzar hablando del proceso para luego ubicar los conceptos en ese proceso.

La actividad de usar datos para desarrollar teorías y modelos y posteriormente usarlos en la práctica es algo a lo que todos estamos habituados. El matiz viene de la integración de ese proceso con determinadas tecnologías para lograr mejores resultados.

Por ello, comencemos desgranando ese proceso en partes suficientemente diferenciadas. Así, estaríamos hablando de:

- La captura y almacenamiento de datos.
- La ordenación y tratamiento de esos datos para sintetizarlos y visualizarlos.
- La búsqueda de relaciones entre los datos, estableciendo constructos matemáticos que relacionan unos datos con otros (funciones, modelos, correlaciones...).
- En base a dichos constructos matemáticos y reglas adicionales, el diseño de algoritmos de decisión que son capaces de predecir eventos, proponer decisiones y tomar acciones.

Dentro de este ámbito es útil distinguir:

- Algoritmos que apoyan la decisión, ofreciéndole ayuda a un operador humano que es el que toma la decisión final
- Algoritmos que están directamente integrados en un proceso automático y toman la decisión de forma autónoma en base al algoritmo definido.

Es importante notar que, salvo en el último apartado (algoritmos integrados en procesos automáticos), no hemos hecho referencia alguna a la necesidad de que intervenga una máquina en el proceso. Todos los subprocesos pueden ser efectuados por una persona incluso con papel

y lápiz. Esto es importante para entender que la ciencia de datos en si misma no requiere de computadoras. Estas hacen más rápido y eficiente el trabajo y, en algunas ocasiones, como veremos más adelante, con procesos que posiblemente no podría replicar un humano. Esta capacidad diferencial es lo que nos hace pensar que la ciencia de datos es algo asociado a la tecnología, pero no necesariamente.

El triaje en los servicios de urgencias

Un ejemplo bastante tradicional y visual de este hecho es el proceso de triaje en las urgencias hospitalarias. Ante una demanda que supera la capacidad del servicio de urgencias, se pone en marcha un proceso que clasifica a los pacientes según la importancia de ser atendidos con verdadera urgencia, a través de una serie de datos y variables recogidas en el proceso de entrada.

Es un proceso esencialmente manual y donde interviene el juicio experto, pero donde al mismo tiempo existen unos protocolos predeterminados en base a la larga experiencia clínica existente y que son el algoritmo de decisión que se facilita al médico para que adopte la decisión final.

No es necesaria una máquina, aunque la incorporación de las capacidades tecnológicas actuales puede mejorar la captura de datos, la identificación de mejoras en el protocolo e incluso su aplicación práctica a los pacientes.

El ajedrez

Un ejemplo que nos da otra perspectiva es el juego del ajedrez. Ya sabemos que existen hoy día computadoras que son capaces de ganar a los mejores jugadores humanos. El proceso no deja de ser el mismo. Podría pensarse que el programa diseñado para jugar al ajedrez usa lo que se denomina “fuerza bruta”, es decir, verificar todos los posibles movimientos alternativos y decidir aquel que le lleva a la victoria. Pero no es cierto, la cantidad de variaciones posibles no hacen viable este camino.

Lo que hacen estas máquinas es calcular, a través de un modelo matemático, que posición de tablero es más ventajosa para sus intereses analizando las posibles alternativas de un numero de jugadas determinadas. Por tanto, usan un cálculo de probabilidad en función del análisis de los datos de millones de partidas y de identificar que situaciones suelen ser más ventajosas para la victoria final.

En este caso, como vemos, el ordenador es capaz de sintetizar y ejecutar un proceso que el maestro de ajedrez realiza intuitivamente en base a su experiencia y conocimiento, pero con una precisión mayor, dado que es capaz de evaluar un número elevadísimo de opciones (aunque no todas) y asignarles una probabilidad específica para la victoria.

De nuevo es el mismo proceso: obtener datos, clasificarlos, crear un modelo matemático y crear un proceso de decisión en base a dicho modelo y otras variables que pueden incorporarse, como, por ejemplo, el tiempo máximo para realizar una jugada y que condicionaría el orden y volumen del análisis.

Definiciones

Inteligencia Artificial

Llegados a este punto, podemos que la Inteligencia Artificial (IA) podría considerarse como el uso de sistemas de computación para ejecutar los procesos anteriormente descritos y que, fundamentalmente, procesan datos, crean modelos y/o ejecutan algoritmos para la predicción, recomendación o decisión de acciones.

Este es el concepto más generalizado de la inteligencia artificial y que es el recogido por las principales normas y recomendaciones de aceptación internacional:

La recomendación sobre ética de la inteligencia artificial de la UNESCO “considera los sistemas de IA sistemas capaces de procesar datos e información de una manera que se asemeja a un comportamiento inteligente, y abarca generalmente aspectos de razonamiento, aprendizaje, percepción, predicción, planificación o control.”¹

La recomendación del Consejo de la OCDE sobre Inteligencia Artificial indica que un Sistema de inteligencia artificial es un sistema basado en una máquina que, para un objetivo definido por un humano, es capaz de hacer predicciones, recomendaciones o decisiones que influyen en el mundo real o virtual. Estos sistemas pueden tener diferentes grados de autonomía.”²

La propuesta de reglamento de la Comisión Europea, en la denominada Ley de la Inteligencia Artificial, define esta como “el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I (*estadísticas, de lógica o de aprendizaje*) y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa.”³

No podemos dejar de mencionar que la inteligencia artificial tiene, además, un componente filosófico y que se refiere a la posibilidad de que las computadoras sean capaces de tener consciencia y sentir, emulando otras características humanas que van más allá de la inteligencia práctica. Para este terreno se suele usar hablar de “inteligencia artificial fuerte” y, en estos momentos, no entra dentro de nuestro ámbito de interés, dado que hoy no existen los conceptos y herramientas para llegar, si es posible, a ese nivel.

Big Data

Hasta este momento no hemos mencionado el big data, pues no era necesario.

Definiremos el big data como la inteligencia artificial cuando, además, aprovecha las capacidades propias del big data y que tienen que ver concretamente con tres aspectos: la capacidad de manejar enormes volúmenes de datos, pudiendo ser estos de naturaleza muy diversa (variedad de datos), lo que incluye imágenes, símbolos, sonidos, ondas y cualquier otro elemento que pueda tratarse y, además, es capaz de hacerlo con una velocidad de proceso suficiente para que el tratamiento sea eficiente. Por tanto, volumen, variedad y velocidad.

Hay autores que han tratado de añadir otras características (siempre buscando que empiecen por “v”, para seguir la serie). En general, consideramos que esas adiciones no son específicas del

¹ UNESCO. Recomendación sobre la ética de la inteligencia artificial. Documento 41C de la 41ª Conferencia General de la Unesco. París, noviembre de 2021 (#N003-01)

² OECD, Recommendation of the council on artificial intelligence, adopted by the Council at Ministerial Level on 22 May 2019. (#N004-01). En inglés en el original: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

³ Consejo Europeo. Propuesta de reglamento del parlamento europeo y del consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la unión (#N007-01). La indicación entre paréntesis en el texto la hemos añadido para clarificar el contenido del anexo citado.

big data, sino que son aplicables a cualquier conjunto de datos (veracidad, viabilidad, visualización y valor).

Robótica

Hablaremos de robótica cuando hablamos de los elementos mecánicos que, de alguna forma, dan vida a los procesos de la inteligencia artificial fuera del software o máquina de computación. Por tanto, es un complemento adicional, de elevada complejidad física, que es capaz de realizar acciones físicas en base a los algoritmos de inteligencia artificial (aunque existe elementos robóticos que son puramente mecánicos, sin intervención de la inteligencia artificial. Un buen ejemplo es la diferencia entre un coche puramente mecánico frente a un coche con diferentes sistemas de ayuda a la conducción).

Constructo matemático (o modelo)

Llamaremos constructo matemático (o modelo) al conjunto de elementos de carácter lógico, matemático, estadístico o relacional que permiten establecer diferentes relaciones entre los datos que se usan para la construcción de dicho modelo y que sirven de base para identificar conclusiones semejantes en los datos nuevos que se introducen en el sistema.

Tratamos de evitar el uso de la palabra modelo al llevar más fácilmente a confusión en la lengua inglesa.

Algoritmo

Reservamos este concepto para el subproceso final que, a partir del modelo o constructo matemático, hace finalmente la predicción, recomendación o decisión. Lo separamos del modelo matemático porque, generalmente, incluye componentes adicionales como restricciones, prioridades o exclusiones en el resultado final que van más allá del propio resultado del modelo.⁴

Aprendizaje automático (machine learning)

Todos los sistemas que hemos descrito pueden evolucionar a medida que conseguimos nuevos datos o mejoramos los constructos matemáticos que los relacionan. Este proceso de mejora lo pueden hacer personas con ayuda de computadoras o, en determinadas circunstancias, lo pueden hacer las computadoras por sí mismas, si bien ese aprendizaje a la vez puede ser supervisado por personas en mayor o menor medida (o incluso no supervisado).

Sin embargo, la distinción anterior, siendo relevante, no es la que nos parece esencial para hablar de aprendizaje automático. Reservaremos este concepto para aquellos casos en que dejamos que el sistema de inteligencia artificial evolucione con cierto grado de autonomía. Es decir, en aquellos casos en que permitimos que el sistema aprenda por sí mismo y que, además, de ese aprendizaje puede auto modular el propio sistema, llegando a cambiar el sentido de sus predicciones, recomendaciones o decisiones en base a dicho aprendizaje, sin que ningún humano intervenga en el proceso a priori.

⁴ Un algoritmo es un conjunto de instrucciones o reglas definidas y no-ambiguas, ordenadas y finitas que permite, típicamente, solucionar un problema, realizar un cómputo, procesar datos y llevar a cabo otras tareas o actividades. De acuerdo con esta definición, el algoritmo puede incluir numerosos constructos matemáticos y otros procesos, incluyendo la toma de datos, pero su esencia está en que arroja un resultado final. Por eso reservamos ese concepto para el subproceso que, usando todos los anteriores, genera la decisión, recomendación o predicción solicitada.

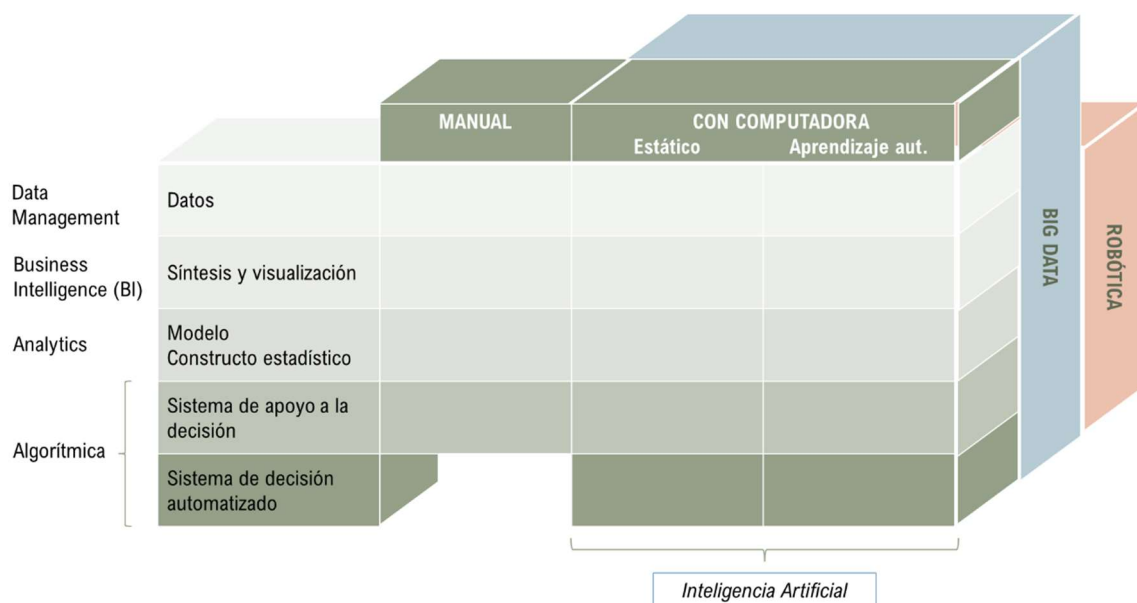
Ubicando el Big Data y la Inteligencia Artificial

Por tanto, insistimos, en nuestra terminología, hablaremos de aprendizaje automático cuando permitimos que el sistema se modifique de forma autónoma, independientemente de cómo llega a ese aprendizaje.

Otros términos relevantes

Volviendo a nuestra descomposición del proceso de desarrollo de algoritmos, podemos decir que el proceso de acopio y almacenamiento de datos lo podemos encajar en el concepto de *data management*; la síntesis y visualización de datos como *business intelligence (BI)*; la construcción de modelos matemáticos sería el *data analytics* y el desarrollo de algoritmos de decisión lo denominaremos *algorítmica*. El conjunto de todos estos subprocesos es lo que denominamos *data science*.

El siguiente esquema representa los conceptos anteriores de una forma visual, a través de los componentes del proceso (primera columna vertical) y de los diferentes sistemas empleados.



Conclusiones

El concepto de inteligencia artificial es muy amplio. Así se quiere ver para evitar que los avances en ciernes en este campo puedan quedarse fuera. Implica cualquier sistema informático que permite la predicción, recomendación o decisión de acciones, basados normalmente en datos y constructos matemáticos de naturaleza muy diversa.

Esto quiere decir que llamamos inteligencia artificial a cualquier predicción, recomendación o decisión en la que interviene una máquina en su realización.

Cuando hablamos de ética algorítmica estamos, precisamente, hablando de la ética de las decisiones y las especiales consideraciones que implica cuando lo realiza un computador. Por tanto, incluye el vasto conocimiento que existe sobre ética de las decisiones, y añade la peculiaridad de la existencia de un sistema automatizado que interviene en el proceso, con sus ventajas y riesgos asociados.