

Febrero 2017

Aprendizaje automático

Extracto de la
Enciclopedia Ipsos

Presentación de la Enciclopedia Ipsos

Nuestra enciclopedia es uno de los primeros proyectos del Knowledge Centre, el nuevo centro de conocimiento de Ipsos; su misión es definir, organizar y compartir el conocimiento que Ipsos ha adquirido a lo largo de numerosos años gracias al trabajo realizado con sus clientes. El objetivo de la enciclopedia es recopilar los mejores conocimientos de Ipsos de forma accesible para todos.

Las definiciones han sido redactadas por algunos de nuestros miembros más expertos y se han sometido a un proceso de revisión externa por especialistas para garantizar que representan la visión más actual de Ipsos. Queremos que nuestra enciclopedia contribuya de forma tangible no solo al trabajo que realizamos, sino también al sector en general.

CONTENIDO

A continuación le mostramos una definición de la primera edición de nuestra enciclopedia. El recurso completo contiene cientos de definiciones y cada una de ellas explica un término concreto relevante para el sector de los estudios de mercado.

Muchas de estas definiciones también incluyen referencias a otras fuentes, tanto de Ipsos como externas. A medida que desarrollábamos el enfoque de trabajo, nos dimos cuenta de que algunos términos tienen una mayor importancia estratégica para el sector y, por lo tanto, era necesario explorarlos en mayor profundidad.

En estos casos, la definición también incluye una sección dedicada al «punto de vista de Ipsos», que expresa nuestras opiniones y perspectivas al respecto, y que se funda en las experiencias y opiniones de nuestros expertos. La definición que leerá a continuación es un claro ejemplo de ello.

¿PARA QUÉ ES Y A QUIÉN ESTÁ DESTINADA?

La enciclopedia ha sido diseñada para asistir a nuestros investigadores y clientes. Su objetivo es ayudarles a comprender mejor cómo trabaja Ipsos, contribuir al desarrollo profesional de nuestro equipo y elevar, aun más, los niveles de calidad.

¿CÓMO SE UTILIZA?

A continuación le mostramos uno de los fragmentos de la enciclopedia de la versión para imprimir. Está íntegramente disponible en formato digital para el personal de Ipsos, y también contiene información sobre las personas que la han redactado. Actualmente, esta versión solo está disponible como recurso interno de Ipsos, pero si desea más información al respecto, póngase en contacto con nosotros: ikc@ipsos.com

EL FUTURO

Al igual que cualquier obra de conocimiento escrita, nuestra enciclopedia nunca se dará por terminada, sino que continuará desarrollándose y creciendo a lo largo del tiempo. Las definiciones se revisan continuamente y se añaden otras nuevas; además, se renueva el grupo de colaboradores. Asimismo hemos creado un proceso de mejoras fundadas en las opiniones, de manera que nuestros lectores puedan proporcionarnos sus pensamientos al respecto y formular preguntas directamente a los expertos que la redactan.

Este mismo principio se aplica a los fragmentos impresos: nos encantaría conocer qué opina, en general, sobre los conceptos que se describen en la enciclopedia y, en particular, sobre el contenido de esta definición. Por ejemplo, ¿cree que es fácil de utilizar y comprender? ¿Cree que falta información que debemos incluir? ¿Ve cómo podría aplicarlo a su trabajo? Póngase en contacto con nosotros: ikc@ipsos.com

Aprendizaje automático

DEFINICIÓN

EL APRENDIZAJE AUTOMÁTICO es un campo de la informática que abarca el diseño y la implementación de algoritmos que optimizan un criterio de rendimiento. Programamos los ordenadores para aprender a partir de ciertas combinaciones de datos de ejemplo (entrenamiento), previo conocimiento del dominio, y con el refuerzo de prueba y error, a fin de predecir, clasificar o agrupar datos nuevos. El aprendizaje automático tiene sus fundamentos en las matemáticas, expresadas en códigos y convertidas en resultados a través de las herramientas.

La tarea básica del aprendizaje automático consiste en extraer conclusiones a partir de una muestra (es decir, generalizar un conjunto limitado de datos). Gracias al desarrollo de programas que mejoran automáticamente con la experiencia, los procedimientos de aprendizaje automático se adaptan al entorno cambiante, y detectan y extrapolan patrones significativos a partir de datos.

Utilidad del aprendizaje automático

Como decía Arthur Samuel, uno de los pioneros de la inteligencia artificial: "El aprendizaje automático es el ámbito de estudio que aporta a los ordenadores la capacidad de aprender sin estar explícitamente programados". En este aspecto preciso es donde resulta más útil el aprendizaje automático; más en concreto, cuando no se puede resolver un problema de manera directa.

Los ordenadores necesitan secuencias de instrucciones para producir resultados a partir de determinados datos de entrada. Sin embargo, hay algunas tareas para las que no resulta evidente un algoritmo o este cambia con

el tiempo, como puede ser el reconocimiento de voz o la detección de operaciones fraudulentas con tarjetas de crédito. Si bien es posible acceder a los datos de entrada y el programador puede saber qué resultados debería generar, no resulta obvio el método para transformar los datos de entrada en resultados.

Al aportar suficientes datos de ejemplo, los algoritmos de aprendizaje automático pueden aprender a identificar el resultado. Esto se debe a que se asume que existe un proceso que explica los datos observados (que existe un proceso que no es totalmente aleatorio) y se pueden extraer los patrones que subyacen en los datos. Aunque no se puede definir de manera directa este mecanismo, sí se puede elaborar una aproximación válida y útil.

Modelo predictivo frente a modelo descriptivo

Los modelos de aprendizaje automático pueden ser predictivos, con el fin de obtener una precisión predictiva en el futuro, o descriptivos, con el fin de adquirir información; o bien una combinación de ambos.

Esta distinción también se puede considerar en cuanto al tipo de aprendizaje: supervisado o sin supervisión. El aprendizaje supervisado es un tipo de técnica de aprendizaje en el que se "entrena" a los algoritmos en un conjunto inicial de datos y, a continuación, se prueban en un conjunto nuevo de datos. El aprendizaje sin supervisión se produce cuando no hay algo en concreto que predecir, sino que el programador simplemente busca patrones que surjan de forma natural de los datos.

Una descripción más útil, aunque también más discutible, de esta distinción en el aprendizaje automático consiste en dividir el espacio en aprendizaje predictivo y aprendizaje por representación. Según esta descripción, en el aprendizaje predictivo se analizan los datos extraídos de alguna distribución y se busca predecir algún aspecto de los valores observables a través de una tarea predictiva bien definida (p. ej., bosques aleatorios o redes neuronales). En el aprendizaje por representación no se predice el diseño, sino que se observa un aspecto de la estructura subyacente, de forma que esta representación pueda ayudar a responder a diversas cuestiones posteriores (p. ej., agrupación, reducción de la dimensionalidad, etc.).

El proceso

En método general de las tareas de aprendizaje automático predictivo consiste en comenzar con una fase de entrenamiento, en la que se inicia el modelo y se “entrena” con los datos de ejemplo. La siguiente fase es la fase de prueba, en la que se utilizan datos nuevos para comprobar la validez del modelo. En esta evaluación, se debe realizar una validación cruzada y medir el rendimiento mediante algunas tasas de error relativas a un criterio optimizado.

En el caso del aprendizaje automático por representación, al no existir datos etiquetados, la tarea consiste en descubrir una representación de datos analizando las similitudes y las diferencias entre los datos y los grupos de datos, extrayendo las características más descriptivas, o encontrando atributos compartidos o clases latentes. En este caso, no existe un método estándar para medir el rendimiento.

La extracción y la selección de características es posiblemente la fase más importante, y a menudo la más ignorada, del aprendizaje automático. Las características de gran valor son mejores que los algoritmos. Es importante disponer de conocimientos sobre un dominio especializado a la hora de seleccionar características. Además, al elaborar el modelo, se debe poner el mayor énfasis en tratar de idear métodos que describan mejor el dominio, al tiempo que se debe adoptar una actitud clara y realista con respecto a las características que se pueden descubrir realmente.

La caja negra

En los modelos predictivos, los algoritmos de aprendizaje automático en ocasiones se denominan “cajas negras”. Esto significa que, en estos casos, los programadores no se suelen centrar en la interpretación de los parámetros o en la explicación del proceso generativo subyacente; y, cuando lo hacen, es con la intención de ajustar manualmente las características o el modelo únicamente para aumentar su poder predictivo.

El aprendizaje automático en contexto

En contraposición al aprendizaje como conducta humana, en el que las habilidades y el conocimiento se desarrollan de forma flexible, y se aplican a una amplia variedad de contextos y problemas, el aprendizaje automático trata de descubrir patrones en la información de entrada que se corresponden con un resultado muy concreto. Por ello, los algoritmos de aprendizaje automático casi siempre requieren un conjunto importante de datos para ajustar el modelo.

El aprendizaje automático se puede considerar pariente de la inteligencia artificial en tanto que se interesa por los sistemas inteligentes que pueden aprender y adaptarse a un entorno cambiante. El modelo se define con algunos parámetros, y la función del aprendizaje es optimizar estos parámetros o mejorar la medición de su rendimiento.

Extracción de datos

La extracción de datos es un concepto que en ocasiones se confunde con el aprendizaje automático. Se trata del proceso de extracción de modelos de datos útiles, normalmente a partir de bases de datos de gran tamaño. Uno de los métodos que se utilizan para descubrir estos modelos es utilizar algoritmos de aprendizaje automático para extraer los datos. El término “extracción” se utiliza debido a su analogía con la extracción de materias primas en minas y el procesamiento de estas a fin de conseguir cantidades reducidas de metales preciosos. En la extracción de datos, los volúmenes de datos se procesan con el fin de elaborar modelos simples y útiles. Otros ejemplos de técnicas de extracción de datos incluyen la síntesis y la extracción de características.

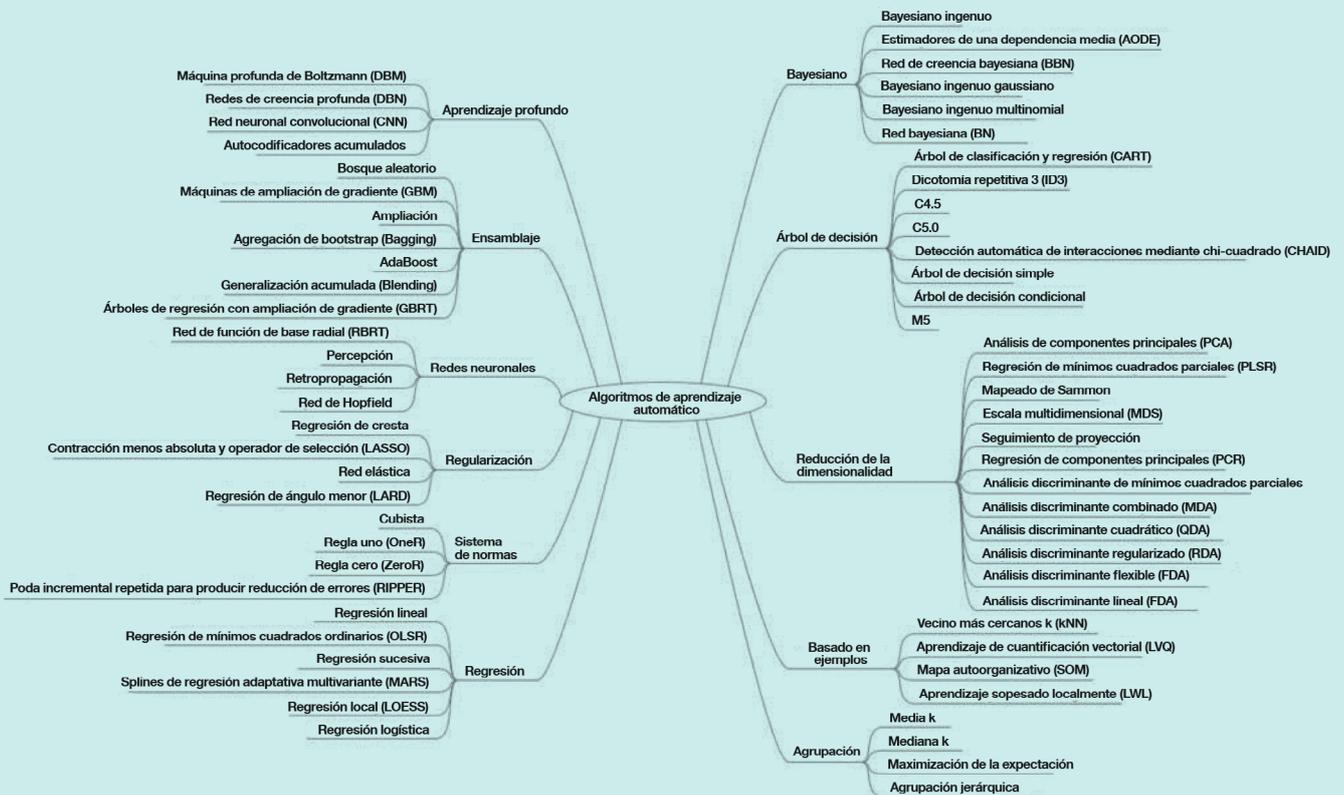
Aprendizaje profundo

El aprendizaje profundo es una vertiente del aprendizaje automático que aprende jerarquías o representaciones de características, que a menudo se basa en redes neuronales artificiales, y que se suele utilizar en ámbitos como el procesamiento del lenguaje natural, la visión computacional y la bioinformática.

Métodos de ensamblaje

En el ámbito del aprendizaje automático, el ensamblaje consiste en utilizar conjuntamente varios métodos para obtener un mayor poder predictivo frente a lo que se podría aprender mediante algoritmos independientes.

Topología del aprendizaje automático



El aprendizaje automático puede tener una definición muy amplia, en el sentido de cualquier método por el que un ordenador puede mejorar su rendimiento adquiriendo experiencia. En un sentido más estricto, o en la literatura popular, el término se utiliza a veces como una palabra exótica de moda y se confunde con la inteligencia artificial a un nivel que realmente no existe en la tecnología actual.

En el caso del análisis de encuestas, y datos transaccionales o conductistas, que son los que normalmente recopila Ipsos, el aprendizaje automático se encuentra a medio camino entre estos dos extremos; se puede considerar una colección diversa de algoritmos avanzados y, en la mayoría de los casos, altamente computacionales. Estos programas suelen implementarse de forma repetitiva, comprobando el rendimiento conforme se implementan y continuando su implementación según sea necesario a fin de mejorar el rendimiento. Con esta definición, no se considerarían aprendizaje automático los métodos que se reducen a una solución formulista relativamente simple.

Así pues, según esta definición, los métodos más tradicionales, como la regresión lineal, no se considerarían aprendizaje automático. Si tiene alguna pregunta a cerca de estas técnicas u otras técnicas específicas, póngase en contacto con su equipo de ciencia de marketing local o con un equipo internacional, como Ipsos Science Centre.

Hay que tener en cuenta que el aprendizaje automático no es una solución mágica para cualquier necesidad. En el caso de que las variables presenten una relación débil o nula con un resultado, poco pueden hacer las técnicas de algoritmos avanzados para solucionar el problema. En estos casos, las principales mejoras para el modelo suelen surgir de una reflexión más profunda sobre las variables que se deben incluir o sobre cómo definir las.

Otra advertencia que debe tenerse en cuenta es que algunas personas piensan inicialmente en las técnicas avanzadas en términos del concepto de “caja negra” que vimos anteriormente. Esto puede llevarles a pensar que no pueden comprender o explicar el análisis a las partes interesadas. Los expertos en ciencia de marketing a menudo pueden facilitar esta comprensión (y, con ello, contribuir a fortalecer la posición de Ipsos como líder de opinión), aunque hay que estar preparado para esta crítica.

Por último, algunos métodos de aprendizaje supervisado generan, junto con sus predicciones, varias medidas de importancia predictiva variable. Suele tratarse de medidas de importancia complejas o académicas, más que de medidas de importancia prácticas, y no se deben utilizar como puntuaciones de análisis de motivaciones ni confundirse con estas.

LECTURAS RECOMENDADAS

Aprendizaje automático para análisis de datos predictivos, algoritmos, ejemplos elaborados y estudios de casos por J.D. Kelleher, J.D. , B.M Namee y A. D’Arcy (2015) The MIT Press, Cambridge, MA.

Obtención de una visión general por C. Ho, Punto de vista de Ipsos (uso de bosques aleatorios)

Introducción al aprendizaje automático por Ethem Alpaydin (2010), The MIT Press, Cambridge, MA.