



# El horizonte de aplicaciones de Inteligencia Artificial en la Ciencia de Datos

**En diálogo con Bit&Byte, el prestigioso Doctor José Ángel Olivas de la Universidad de Castilla - La Mancha de España se refirió al avance en la Ciencia de Datos. Durante la conversación destacó que el desarrollo de sistemas inteligentes es “el cambio de paradigma social más grande de la historia” y que la industria informática necesita de más profesionales con una formación abarcativa.**

## **-¿Cómo imagina el futuro de las aplicaciones de Inteligencia Artificial en la Ciencia de Datos?**

Yo creo que la ciencia de datos y el análisis de datos pasa necesariamente por el uso de ciencias sofisticadas de Inteligencia Artificial.

La Inteligencia Artificial clásica no está dando buenos resultados para el análisis inteligente de datos. De hecho, las aplicaciones de predicción que funcionan bien son las que se basan en la estadística. Sin embargo las de pronóstico, por ejemplo, terremotos, puntos de inflexión o una problemática económica actual, como cuándo va a parar de subir el dólar, esas parece que no funcionan bien.

La Inteligencia Artificial desde mi punto de vista está estancada en los paradigmas clásicos de redes neuronales, lógica borrosa, algoritmos

genéticos, clustering basado en estadística, etc. Y está tan estancado que la Inteligencia Artificial tiene que proponer nuevos modelos adaptados al nuevo mundo con las cantidades ingentes de datos que hay para que sean capaces de tomar medianamente inteligente esos datos.

Entonces yo soy bastante escéptico, creo que hay que hacer mucho, creo que la disrupción tecnológica de la que se habla ahora es un mito en la sociedad, se piensa que con los datos todo el mundo sabe todo de nosotros y eso es mentira. La mayoría de las empresas no tienen ni idea de qué hacer con los datos, son muy pocas las que tienen la capacidad de hacer cosas serias con los datos, en general son las empresas grandes desde el Silicon Valley, Google, Facebook, etc.

## **-¿Ud. entiende que el avance en la Ciencia de Datos tiene/tendrá un**

## **impacto significativo en la industria del software y en las industrias TIC en general? ¿Por qué?**

Creo que la Ciencia de Datos como tal no existe, es decir, hay datos, información, conocimiento con los que hay que hacer algo. Una cosa es almacenarlos inteligentemente, otra cosa es almacenarlos, otra es transmitirlos y otra diferente es pre procesarlos.

Claramente no es que tenga un impacto significativo en la industria del software, es que lo determinará porque determina el presente y el futuro. Hoy, el 90% de la industria del software está dedicada a almacenamiento, gestión y análisis inteligente de datos. Por lo tanto, es una paranoia actual de la sociedad –mal llamada del conocimiento– porque ahora estamos en la sociedad de los datos.

Los datos se manejan por sensores, los teléfonos celulares llevan sensores

que pueden medir 20 magnitudes por segundo y todo eso ¿Dónde se guarda? ¿Cómo se utiliza? Yo creo que la sociedad no sólo de software sino la sociedad tecnológica en gran medida, está orientada a la capacidad de absorber y utilizar los datos que se están generando.

### **-¿Qué relación existe entre la Inteligencia Artificial en general con los datos masivos?**

Los datos se generan de forma independiente a que vayan a ser analizados o no. Es decir, cuando yo soy un cliente de un servicio de telefonía, mis datos se están generando porque yo compré un teléfono y tienen que guardar las llamadas que hago, a qué hora las realizo, a dónde me comunico. Esos datos están guardados independientemente de que se usen en forma inteligente.

Hay otras disciplinas, por ejemplo, las teorías de las bases de datos que están relacionadas con el almacenamiento. La Inteligencia Artificial tiene que ver con el análisis (contemplado de forma general porque tiene que ver con machine learning y algunas ciencias sociales) y es la única forma de analizar inteligentemente los datos. Lo que pasa es que se avanza despacio, lo que se hace es muy básico y no se está haciendo bien, es un tema complicado que se va a resolver a largo plazo; nosotros no lo vamos a ver.

Hay un problema muy duro que es el procesamiento del lenguaje natural, los humanos no vamos a dejar fácilmente que un engendro automático entienda todos los matices que decimos, nuestros sentimientos, etc. Se puede por ejemplo hablar con un sistema que te permita decidir entre opciones, pero no detectar el sentido, la ironía, el contexto en que se dice algo determinado.

Necesitamos una larga trayectoria y la Inteligencia Artificial con sus asociados: Psicología cognitiva, Sociología, Lingüística, Estadística. Si se aprende a manejar bien todo junto, se podrán diseñar algoritmos y mecanismos que permitan analizar datos de forma más inteligente.

### **-¿Qué consecuencias considera que trae el desarrollo de sistemas inteligentes a nivel social y económico?**

- Están cambiando la sociedad, el concepto de sociedad se modificó radicalmente. Creo que es el cambio



de paradigma social más grande de la historia. Hoy una serie de comentarios negativos de un blog sobre un determinado producto puede hundir una empresa; esto antes era imposible.

Además, por ejemplo el comercio de turismo y viajes con empresas como Amazon. Las agencias de turismo desaparecerán porque ya nadie prácticamente compra un vuelo de esa manera. Lo mismo ocurre con la música, ya casi no hay soportes físicos. El modelo económico de empresa tradicional de los últimos dos siglos desaparece porque hoy puede ocurrir que una planta robotizada con dos o tres personas haga lo que antes se hacía con 4 mil. Además son tres personas que no saben de ese negocio, saben de ingeniería en sistemas o de Inteligencia Artificial.

El modelo social y económico está siendo transformado de forma radical, de hecho hoy hay una gran cantidad de gente que trabaja desde la casa, esto hace 10 años era impensado.

Como todo, tiene sus matices positivos y negativos. Como positivo creo que a lo mejor hay más tiempo y menos esfuerzo físico, pero a su vez también trae como negativo que quien no se adapta a la

tecnología queda afuera y tiene muchos problemas. Sin embargo creo que no hay que mirarlo desde este punto de vista de negativo o positivo, las sociedades avanzan y hay que adaptarse, da igual que lo consideres bueno o malo.

### **-¿Cuáles son las técnicas computacionales inteligentes que posibilitan un mejor aprovechamiento de la información almacenada en grandes bases de datos?**

En general hay dos grandes grupos de técnicas de aprendizaje a partir de datos. Los datos en sí mismos se pueden ver como elementos individuales sobre una realidad. Es una instancia, un ejemplo de un determinado problema. Es decir, si yo tengo datos de medicina, cada registro, cada dato de un paciente será sobre sus síntomas, por ejemplo.

Los mecanismos que generalizan, es decir que buscan patrones a partir de datos individuales, es lo que se conoce como razonamiento inductivo, de casos particulares a lo general. Por lo tanto, cualquier algoritmo de aprendizaje que trate de encontrar patrones a partir de datos masivos, necesariamente tendrá

un componente inductivo. De hecho, los árboles de decisión y muchas redes neuronales son algoritmos inductivos.

Dentro del concepto de razonamiento por inducción, de tratar de generar un patrón a partir de casos particulares, tenemos dos grandes grupos:

Por un lado los de clasificación pura, que se llama aprendizaje supervisado, por ejemplo cuando se dice que una persona tuvo determinados síntomas y luego tuvo gripe. Busca las causas y las consecuencias para el diseño de un algoritmo que las relacione.

Por otro lado están los algoritmos de clustering, de aprendizaje no supervisado, que es cuando existen masas de datos pero no tienen clase a la que pertenecen, en nuestro ejemplo sería la gripe. Es decir, yo tengo un montón de páginas web pero sólo puedo agruparlas por sus similitudes, agrupa un conjunto de datos en base a su parecido.

Dentro de estos dos grandes grupos, lenguaje supervisado (clasificación) y el no supervisado (clustering) cabe prácticamente todo: las técnicas estadísticas, la extrapolación, etc.

Estas son las técnicas que se están utilizando, tengo mis dudas acerca de si son suficientes. Ahora hay un nuevo paradigma del cual yo era detractor pero creo que es el único que está funcionando y que es la única salida, consiste en utilizar el componente estocástico de la información.

Es decir, cuando hablamos de modelos muy conocidos como el deep learning, se utiliza la fuerza bruta computacional para poder probar generar aleatoriamente modelos e ir quedándose con los mejores. De esto se trata aprovechar el componente estocástico de la información y creo que cubre todo lo que tiene que ver con la Inteligencia Artificial y el análisis de datos.

**-Como institución académica ¿Qué aspectos considera que debe tener en cuenta la Facultad a la hora de formar profesionales capacitados en Sistemas Inteligentes orientados al procesamiento de Big Data?**

Creo que debe pensarse en una formación amplia en todos los aspectos que hemos hablado. Muchas veces hay un gran error en la mayoría de los masters

en Big Data, en las especializaciones y postgrados que sólo se orientan a herramientas comerciales, librerías, etc.

Eso es lo que no debe hacerse.

Como puede observarse, se necesitan profesionales con una visión global, con conocimientos básicos de Inteligencia Artificial, de Psicología cognitiva, de Sociología, sobre bases de datos y todo orientado al mismo fin.

Pienso que el papel clásico de un informático puro y determinista, desarrollador, programador, tiene probabilidades de éxito nulas analizando datos.

Se requiere una perspectiva más amplia, saber que técnicas son más adecuadas para cada problema y esto tiene que ser tenido en cuenta por las instituciones académicas. Como es el caso de la especialización que se está impartiendo en la Facultad de Informática de la UNLP en Big Data, donde tratamos de dar esta visión panorámica, no sólo algunas técnicas que permitan casi de una forma ciega analizar datos, hay que saber lo que se está haciendo.

Se deben considerar estos aspectos diferentes y más amplios y no sólo programación y herramientas, porque eso es necesario que se conozca, pero no es todo. Una persona que no sepa nada de cómo se organizan los datos, de la sociedad, de inducción, se le puede enseñar cualquier librería y podrá operar muy bien con el ordenador, pero no sabrá hacia dónde va ni qué algoritmo va y por qué.

Entonces no sólo debe haber una formación en los aspectos tecnológicos del manejo de datos y computacionales, es vital la formación en los aspectos cognitivos ■